DOCUMENT RESUME

ED 181 714 FL 010 655

AUTHOR	Simons, Gary Francis
TITLE	Language Variation and Limits to Communication. Technical Report No. 3.
INSTITUTION	Cornell Univ., Ithaca, N.Y. Dept. of Modern Languages and Linguistics.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	79
GRANT	NSF-BNS-76-06031
NOTE	234p.; Revised version of Ph.D. dissertation, Cornell University
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Communication (Thought Transfer); Contrastive Linguistics; Culture Contact; Diglossia; Educational Needs; Field Studies; Language Planning; Language Programs; *Language Research; Language Tests; *Language Variation; Models; Mutual Intelligibility; Needs Assessment; Prediction; *Regional Dialects; *Research Methodology; *Sociolinguistics; *Speech Communication; Vocabulary
IDENTIFIERS	*Solomon Islands

ABSTRACT

Strategies are developed for understanding how language variation limits communication. Methods of measuring communication are discussed, including an intelligibility measure used in the Solomon Islands. The analysis of data gathered using communication measurement is discussed. The result of the analysis is a determination of the number of vernacular language programs needed in a given area, and the proper location for those programs. The measurement of communication and the methodology for finding centers of communication lead to a proposed model of communication. This model suggests that interdialectal understanding depends on the linguistic similarity of two dialects and on the the social relationships between speakers of the dialects. Data from ten field studies on linguistic similarity are analyzed to explore the relationship between lexical similarity and intelligibility. A model for expressing this relationship is proposed. Social data from Santa Cruz in the Solomon Islands are considered in the formulation of a more comprehensive model embracing social and linguistic relationships. Predictions from this model are demonstrated to be over 90% accurate. A bibliography is appended. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181714

LANGUAGE VARIATION AND LIMITS TO COMMUNICATION

Gary F. Simons

Technical Report No. 3

Report to the
National Science Foundation
on Grant BNS 76-06031

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Gary Francis
Simons
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Department of Modern Languages
and Linguistics
Cornell University

Ithaca, New York 14853

FL010655

Technical Reports to NSF
from the Cornell
Department of Modern Languages
and Linguistics

No. 1 Joseph E. Grimes. The Thread of Discourse (1972). Out of print: revision published by Mouton & Co., 1975.

No. 2 Joseph E. Grimes. Word Lists and Languages (1974). Out of print: available as ERIC microfiche ED 100 147.

No. 3 Gary F. Simons. Language Variation and Limits to Communication (1979).

© Gary Francis Simons 1979

PREFACE

This report is essentially a reformatting of my doctoral thesis, presented at Cornell University in January of 1979. The research that went into it, as well as its publication and mailing, were made possible by a grant from the National Science Foundation.

There are many people who have been instrumental in helping me complete this project. I would especially like to thank the members of my graduate special committee, Joseph E. Grimes, Gordon M. Messing, and Ray Teitelbaum. Dr. Grimes, the chairman of the committee, worked closely with me throughout my studies at Cornell, and I owe him much for his guidance. I am grateful for the opportunities he opened up by sending me out as Graduate Research Assistant for the "Language Variation and Limits to Communication" project of which he is Principal Investigator.

I would also like to thank the people who gave technical support for the microcomputers I used in the research. William Hemsath of the Psychology Department at Cornell designed and built the original microcomputer which I used on the field. Ramond Howell and others of the Jungle Aviation and Radio Service implemented a later version of the machine which I used to produce this manuscript.

Several friends willingly offered their time and knowledge to help me in areas which were unfamiliar to me, particularly Gene Chase and Michael Wheeler, who helped with statistical and mathematical problems which were beyond the reach of my training as a linguist. Raymond Gordon, a fellow student in linguistics, has been a constant friend throughout my years at Cornell and has helped me many times by talking through the problems I encountered in research and writing.

My wife, Linda, helped me in immeasurable ways. She served in our project as my assistant in the field research and as project secretary after our return from the field. This means that she handled all the typing of the original manuscript and all the machine editing of subsequent drafts, as well as reading everything and offering comments.

Finally there are the many people who helped to make the research in the field a success. I am indebted to Karl Franklin, Bruce Hooley, and Richard Loving of the Summer Institute of Linguistics in Papua New Guinea for making the arrangements for me to conduct a workshop at their field headquarters in December of 1976. From that workshop resulted a volume of papers (Loving and Simons 1977) and many insights which guided the research in the Solomon Islands the following year. In the Solomon Islands I am indebted to Hugh Paia, Permanent Secretary to the Minister of Education and Cultural Affairs, for making the arrangements with local government councils to allow me to do the research. I am also grateful to the Translation Committee of the Solomon Islands Christian Association and its secretary, the Rev. Robert Stringer, for sponsoring our projects. I must also thank Richard Buchan for assisting me in the dialect survey of Santa Cruz Island, and his wife for extending her hospitality.

Most of all I remember my Solomon Island and Papua New Guinea friends who were helpful and patient with a curious waetman who wanted to study their language. It is impossible to name them all but without their help this work never would have been possible.

Gary Simons
15 November 1978
Butternut, MI

TABLE OF CONTENTS

PREFACE iii

TABLE OF CONTENTS v

LIST OF FIGURES vi

LIST OF TABLES viii

1. INTRODUCTION 1
1.1 An overview 1
1.2 Some definitions: intelligibility and dialect 2

2. MEASURING COMMUNICATION 5
2.1 Conducting an intelligibility survey 5
2.2 A review of intelligibility testing methods 12
2.3 A taxonomy and evaluation of intelligibility testing methods 18

3. FINDING PATTERNS OF COMMUNICATION 33
3.1 The criteria of adequacy and least cost 33
3.2 An inspection method for analyzing patterns of communication 34
3.3 An algorithm for finding all possible least cost dialect groupings 41
3.4 Grimes's optimization algorithm for grouping dialects 47

4. EXPLAINING COMMUNICATION: A MODELING APPROACH 57
4.1 Why build models? 57
4.2 The state of the art 59
4.3 A basic model for explaining communication 62
4.4 Some statistical preliminaries 63

5. EXPLAINING COMMUNICATION: LINGUISTIC FACTORS 67
5.1 Quantifying linguistic similarity 67
5.2 Lexical similarity and intelligibility 69

6. EXPLAINING COMMUNICATION: SOCIAL FACTORS 101
6.1 The characteristics of a dialect system 101
6.2 A general model for explaining communication 115
6.3 Explaining communication on Santa Cruz Island 120

APPENDIX 1 COMPLETE DATA FOR THE STUDY OF SIMILARITY AND INTELLIGIBILITY 143

APPENDIX 2 COMPLETE DATA FOR THE STUDY OF INTELLIGIBILITY ON SANTA CRUZ ISLAND 189

BIBLIOGRAPHY 225

LIST OF FIGURES

2.1 Classification of intelligibility testing methods 20

3.1 Intelligibility on Santa Isabel Island 35
3.2 Patterns of communication on Santa Isabel 36
3.3 The Maringe language program 36
3.4 Three vernacular language programs for Santa Isabel 37
3.5 Intelligibility in Northern Mixteco 38
3.6 Dialect groupings in Northern Mixteco 39
3.7 Hypothetical data 40
3.8 Least cost grouping algorithm 42
3.9 Grimes's optimization algorithm 49
3.10 Threshold corresponds to differences, not actual cost 50
3.11 Threshold corresponds to sums, not individual costs 52
3.12 Dialect groupings on Santa Cruz Island 54

4.1 Scattergram and regression line 64

5.1 Plots for full raw data 72
5.2 Statistics for full raw data 73
5.3 Plots for raw intelligibility with exclusions 77
5.4 Statistics for raw intelligibility with exclusions 78
5.5 Distribution of hometown scores 79
5.6 Plots for adjusted intelligibility with exclusions 82
5.7 Statistics for adjusted intelligibility with exclusions 84
5.8 Linear model 88
5.9 Simplified linear model 89
5.10 Trend line 91
5.11 Step function 92
5.12 Polygonal model 93
5.13 Parabolic model 95
5.14 Constrained parabolic model 96
5.15 Exponential model 97
5.16 Modified exponential model 98
5.17 Logistic model 99

6.1 Relative distance on Santa Cruz Island 112
6.2 Indirect contact on Santa Cruz Island 113
6.3 A linear function and a step function 123
6.4 Single variable predictors of intelligibility 125
6.5 Absolute, relative, and composite distance 126
6.6 Measured contact 129
6.7 Absolute versus relative geographic distance 131
6.8 Absolute versus relative lexical distance 131
6.9 No distance versus relative lexical distance 132

List of Figures - continued

6.10 Composite versus single relative distance 134
6.11 Attraction and motivation 135
6.12 Measured contact versus predicted contact 135
6.13 Direct predictions versus indirect predictions 136
6.14 Summary of successive refinements to models for explaining communication 138

LIST OF TABLES

1.1 Slope 176
1.2 Intelligibility intercept 176
1.3 Percentage of explained variation 177
1.4 Improvement over raw intelligibility 177
1.5 Predicted intelligibility for 100% intelligibility 178
1.6 Deviation from 100% intelligibility 178
1.7 Percentage of similarity for predicted 0% intelligibility 179
1.8 Deviation from the mean value of 40.8% 179

2.1 Geographic distance 191
2.2 Relative geographic distance 192
2.3 Lexical similarity 194
2.4 Lexical distance 195
2.5 Relative lexical distance 196
2.6 Measured intelligibility 197
2.7 Opinions about intelligibility 198
2.8 Attendance at church festivals 200
2.9 Marriage ties 202
2.10 Predicted marital residence 203
2.11 Estimated intelligibility 205

CHAPTER ONE

INTRODUCTION

Language variation limits communication. For this reason, language variation is a vital concern to educators, government officials, broadcasters, publishers, writers, missionaries -- to anyone who has a message to communicate. Many of the developing nations of the world face the challenge of trying to communicate with a multilingual population, a population which may include well over a hundred dialects or languages. Even among nations where a single national language is firmly established, gross dialect variations of the national language and pockets of minority languages still exist.

It may not be thought feasible for a country to initiate projects such as mass communication, bilingual education, or vernacular literature production in every one of its languages and dialects. On the other hand, if that country wishes to reach all of its citizens, it must carry out its programs in languages that are both understood and accepted by all groups concerned. The urgent need then, is for a way to determine which specific dialect or dialects are the most useful in reaching a given population. This thesis develops strategies for understanding how language variation limits communication and for devising solutions which will help overcome these limits to communication.

1.1 An overview

Chapter 2 deals with gathering the fundamental data for a study of language variation and limits to communication. It addresses the question of how to measure communication. It can be measured by devising tests which allow the investigator to observe how well one group understands the speech of another. First I describe in some detail a method of testing understanding which I used in field studies in the Solomon Islands. Then I briefly review a number of methods which other investigators have used. Finally I propose a taxonomy of intelligibility testing methods. My conclusion is that no one method of testing intelligibility is inherently better than another; rather the choice of a method depends on the particular situation. The resulting discussion should serve as a guide to the prospective field investigator for helping select a method of measuring communication which is best suited to his goals and the capabilities of the people among whom he will do the testing.

Fortunately, communicating with every citizen in a particular region does not usually require that a vernacular language program be initiated in each one of its dialects. Chapter 3 tells how the data gathered by the methods of Chapter 2 can be analyzed to determine how many vernacular language programs are needed in an area and where those programs should be centered. A major deterrent to vernacular language programs is the high cost of setting them up and keeping them going. The techniques presented in Chapter 3 find the least costly solutions to establishing vernacular language programs in an area by finding groupings of the dialects which minimize the number of language programs required while at the same time guaranteeing that all citizens will adequately understand the language of at least one of the programs.

Chapters 4, 5, and 6 form a unit on the topic of explaining communication. The methods for measuring communication discussed in Chapter 2 tell us only whether or not communication can take place and to what extent. The methods for finding centers of communication in Chapter 3 allow us to take advantage of measured patterns of communication in finding the least costly solutions to communicating with all the citizens of a region. However, neither method explains why there is communication at all or why the patterns of communication should be what they are. By understanding why patterns of communication are what they are, and not just what they are, it is possible to make better proposals about language planning in an area. Furthermore, by understanding the factors which contribute to intelligibility in an area, it is possible to estimate intelligibility relations which it is not feasible to measure.

The approach to explaining communication is one of building models. Chapter 4 concentrates on the subject of modeling itself. After a discussion of the meaning and advantages of modeling, a basic model for explaining communication is proposed. The model suggests that the amount of understanding between dialects depends on two factors: the linguistic similarity between dialects and the social relationships between them.

In Chapter 5 the factor of linguistic similarity is considered in detail. After a general discussion of various aspects of linguistic similarity and how they can be measured, data from ten different field studies are analyzed to explore the relationship between lexical similarity and intelligibility. As a conclusion, a general model for expressing this relationship is proposed.

In Chapter 6 the second factor of the model, social relations, is considered in detail. First the role of social relations in explaining communication and ways of measuring social relations are discussed. Then data from the island of Santa Cruz, Solomon Islands, are considered. A more comprehensive model which embraces social relationships as well as linguistic ones is used to explain communication between dialects on the island. The predictions derived from this model are over 90% accurate.

1.2 Some definitions: intelligibility and dialect

Before proceeding with the text, two terms need to be defined: intelligibility and dialect. The problem is not so much that people do not know what they mean, but that they mean different things to different people. Therefore, I now define them in the way that they will be used throughout the thesis.

Intelligibility is synonymous with understanding and comprehension. (The root word is *intelligible*, not *intelligence*.) Dialect intelligibility refers specifically to the degree to which speakers of one dialect understand the speech of another dialect. Some linguists who have studied dialect intelligibility restrict the term to mean only a theoretical expected degree of understanding of individuals who have had no experience with the other dialect. For instance, Gillian Sankoff defines intelligibility in this way (1969:839-840). If understanding is boosted by experience with the other dialect, then she contrasts that with intelligibility by calling it "bilingualism". She uses the term "incipient bilingualism" to refer to a degree of bilingualism which does not imply a great deal of learning.

I do not define intelligibility in this way. If a person understands another dialect, then that dialect is intelligible to him. Bilingualism and incipient bilingualism do not contrast with intelligibility; they are special cases of intelligibility. Whenever I refer to that special case of intelligibility which is the theoretical degree of understanding between dialects whose speakers have had no contact, I use the term *inherent intelligibility*.

Another common use of the term in the literature is in the phrase *mutual intelligibility*. This phrase was coined in the early studies of intelligibility in the fifties (Section 2.2.1). Those investigators were actually trying to measure inherent intelligibilty and they averaged the intelligibility in both directions between a pair of dialects in order to approximate a measure of linguistic similarity which they thought should be symmetric. This relationship they termed mutual intelligibility. Somehow the phrase "mutual intelligibility" became interchangeable with the term "intelligibility" in the general literature. They are not interchangeable, however. Intelligibility is not usually a two-way phenomenon. A's intelligibility of B's speech is a different thing than B's intelligibility of A's speech. Intelligibility is mutual if and only if the degree of understanding is the same in both directions. It sometimes is, but asymmetric linguistic and social relations often make it otherwise. Mutual intelligibility is not synonymous with intelligibility; it is another special case of intelligibility.

The second term that needs defining is *dialect*. Two popular level notions of dialect are that it refers to a funny way of speaking or to a way of speaking that differs from a standard or prestigious language. But in a linguistic view the term carries no such connotations; it refers simply to a variety of speech. Some linguists have attempted to define dialect precisely so as to assign it an exact place within a hierarachy of linguistic taxonomy. All such definitions end up being arbitrary, however, and none has received widespread acceptance. The only satisfactory definitions seem to be loose ones. Charles Hockett gives a good example (1958:322):

> A language ... is a collection of more or less similar idiolects. A dialect is just the same thing, with this difference: when both terms are used in a single discussion, the degree of similarity of the idiolects in a single dialect is presumed to be greater than that of all the idiolects in the language.

Throughout this thesis, when I use the term dialect, I shall be referring to a collection of similar idiolects. I use the term *dialect group*, or sometimes just dialect for short, to refer to the group of people who speak those idiolects.

The kinds of dialects which I investigate in this thesis, and which other investigators whom I cite have investigated, are regional or community dialects, that is, the variety of speech which is common to the individuals in a region or a local community like a town or village. Social dialects which cut through regions or communities have yet to be investigated using dialect intelligibility methodologies. Therefore the local community actually serves as the minimal unit in defining the dialects considered in this thesis. That is, dialect refers to the variety of speech common to a local community or a more inclusive grouping of communities. Two dialects are distinguished if their respective speakers recognize that the varieties of speech are different. The degree of difference is not at issue in distinguishing dialects, only the fact that there is a difference.

CHAPTER 2

MEASURING COMMUNICATION

It is a popular notion that people who understand each others' speech speak the same language and people who cannot understand each other speak different languages (Hockett 1958:322, de Saussure 1959:203, Ivic 1974:696). Thus it is that intelligibility testing derives its importance as a method for determining whether or not two different speech communities use the same language (Voegelin and Harris 1951, Wurm and Laycock 1962). Since intelligibility testing was first described in a 1951 article by C. F. Voegelin and Zellig Harris the method has been refined by a number of investigators. Thus far it has reached its fullest development in Eugene Casad's 1974 manual, Dialect Intelligibility Testing.

In this chapter many different methods of testing intelligibility are presented. I begin in Section 2.1 with a detailed discussion of how to conduct an intelligibility survey based on the method of intelligibility testing I used in the Solomon Islands. This method is applicable in situations where the investigator shares a common language with illiterate test subjects. Where subjects are monolingual or literate different methods are appropriate. In Section 2.2 other methods of testing intelligibility are reviewed. The basic outline of conducting a survey remains the same, only the details about constructing and administering the tests differ.

As a conclusion, Section 2.3 develops a taxonomy of intelligibility testing methods and evaluates the situations in which each method is most appropriate. It is argued that no method is inherently better than another; rather, the evaluation depends on the situation in which the testing is done. An optimal method is defined as one which yields the greatest amount of information with the least amount of effort. It is shown that different methods are optimal in different situations. The analysis in Section 2.3 should serve as a guide to field investigators for selecting a method of intelligibility testing.

2.1 Conducting an intelligibility survey

Intelligibility between dialects is measured by observing how well speakers of one dialect understand a recorded text from another dialect. To carry out this testing requires that each dialect area be visited twice, the first time to collect the texts, and the second time to do the testing. An intelligibility survey consists of four steps: (1) planning the survey, (2) collecting the texts, (3) preparing test tapes, and (4) administering the tests. The final step of processing and interpreting the results is treated in following chapters. This basic outline of an intelligibility survey holds for all the methods described in this chapter. The specific details of collecting, preparing, and administering tests describe the method I used in the Solomon Islands (Simons 1977a). For other complete overviews of a dialect intelligibility survey see Linda Simons 1977 and chapter 2 of Casad 1974.

2.1.1 Planning the survey

The purpose of the planning stage is to determine which villages must be visited during the intelligibility testing survey. It is usually not necessary to conduct a test in every village within the survey area. Rather, we need only to test one representative village for each different dialect in the area. Therefore, it is wise to use any maps, census data, or linguistic and anthropological publications about the area to determine the location and extent of each of the different dialect groups within the area. Often there will be very little such material available and the investigator may have to rely almost entirely upon his first visit into the area to gather this information. In this case the information is gathered by talking to local people to gain their opinions about the dialect groupings within the area. Through this questioning the investigator gains a rough sketch of the dialect situation within the area. This preliminary picture is bound to be incomplete. The investigator must be sure to maintain a flexibility to follow new leads as they are uncovered at later stages in the survey.

After all the presumed dialects have been located the investigator can plan a route for the survey trip through the area. Ideally, he should plan to visit one village for each of the dialect groups turned up in this preliminary stage of the survey. The actual villages which are visited may be determined on the basis of the presence of roads or trails or nearness to other villages which must be visited. Local opinions about which villages are important ones should also be considered. It is generally wise to visit the most remote village in the survey area last. If it is visited last there is no need to return again. The test tapes for all the other dialects will have been collected and prepared by that time and the administering of the intelligibility tests can begin at that village.

Another important aspect of the planning stage is a pilot survey in which the methods of collecting, preparing, and administering the tests are tried out in one or two villages before the actual collection phase begins. This trial run may point to modifications needed in the method before it is too late to change.

2.1.2 Collecting the texts

On the first trip through the survey area the investigator stops at each of the villages selected in the planning stage in order to collect texts which are to be used in intelligibility testing. If a more extensive language survey is being conducted, one which also includes study of linguistic similarity and social relations between dialects, these data should also be collected during this first trip through the area. This allows the investigator to have a good look over all the data before making the second trip. During the second trip he will then be more aware of the whole setting and will have opportunity to ask further questions about social relations or to check up on linguistic data that may look questionable.

The informants chosen to give the texts for the intelligibility tests should be native speakers of the local dialect and also speakers of a language shared by the investigator or his assistant. The investigator should first carefully screen the informant to be sure he or she is adequate. This is done by asking the informant where he was born, where his parents were born, if he has lived or worked in any other areas, the languages his parents and spouse speak, and other questions which will help to determine if the informant is truly a native speaker of the local dialect. Special care must be taken in areas where men or women marry into villages other than their own -- half the adults in a village may not be native to it.

It is important that the investigator clearly explain to his informant what kind of text is wanted. The text should be fairly short; an ideal length is two and a half to three minutes, though texts as short as one and a half minutes or as long as five minutes have been used successfully. The subject matter should be autobiographical in nature, rather than folkloristic or procedural. Folkloristic and procedural texts often contain a specialized style or vocabulary. Also, there is a general widespread knowledge of both folklore and procedures which make them unacceptable subject matter for intelligibility testing, because only minimal cues to the content are needed to make all the rest accessible. Thus an autobiographical text which will be unpredictable to the listener in its content is most desirable.

It is helpful for the investigator to suggest topics to the informant. Some possible topics are: what he did yesterday, a favorite hunting or fishing story, a family emergency, or a recent trip. If the investigator already has collected a few good texts from other villages which this informant may understand, it may be helpful to play these for him so he may get an idea of what is expected of him.

The informant may appreciate a practice run to tell his story before it is recorded. This may help to put him at ease, allow him to organize his thoughts, and also give the investigator an idea whether or not the story is appropriate. The investigator can then ask questions about the content and help the informant bring out details in the episode which may improve the quality of the test. If a text is recorded and then proves to be too short, the same kind of technique can be used. The investigator can ask questions about what has been recorded and offer suggestions as to how the text could be expanded. Then the informant can be given a chance to add more to the end of what has already been recorded.

After a good text has been recorded it must be translated into a language which the investigator can understand. This would ordinarily be a trade language or the national language if he is not familiar with the vernaculars in the area. This is best done in an interlinear fashion using two tape recorders (Voegelin and Harris 1951:328, L. Simons 1977:240). The first tape recorder is used to play back the original text in short sections. These sections should correspond to natural breaks in the text. After each section, the storyteller is asked to give a translation of that section. The second tape recorder is left running during this whole process in order to record both the original text and its translation. The result is like an interlinear translation of the original text. The completeness and accuracy of the translation can be verified by getting another translation of the story from someone else or by administering the completed test tape to other speakers of that dialect.

2.1.3 Preparing the test tapes

The first step in preparing a test tape is to transcribe the interlinear translation tape. Unless the vernacular texts are also needed for grammatical analysis or comparison, there is no need to make an exact morpheme by morpheme transcription and translation of the text. The vernacular portion of the text may be transcribed in broad outline only, noting mainly the intonation contours and the final syllables preceding pauses. The translation, however, should be transcribed in full. The complete translation is then studied to break up the text into logical segments. When possible, these segments should be defined both in terms of their content and of having final intonation contours. They should be long enough so that questions can be asked about the content of the segment, but not so long that a listener would be likely to forget what took place at the beginning of a segment before he reached the end. Around fifteen seconds is an optimal length for a segment.

The actual test tape consists of two parts. In the first part, the first one or two minutes of the text are copied without a break. In the second part the entire text is copied in the short segments defined above. The purpose of the first part of the test tape is simply to allow the listeners to tune in to the speaker's voice and to the new dialect which they are about to be tested on. The second part of the test tape which is divided into sections comprises the actual test. In the testing situation this form of the text is played back segment by segment, and after each segment listeners are asked to make a response.

The test tape is made using two tape recorders. In one, the original vernacular text is placed; in the other, a blank tape which will be the test tape is placed. To record the first part of the test, the uninterrupted section of text, the transcription should be studied to find a logical breaking point which is one to two minutes into the text. If the text is short, this first part of the test may include the whole text. If the text is long, it will save time in the testing to cut the text short for the first part. The blank tape is then set to record while the original text is played and this first section is dubbed onto the blank tape. At the selected breaking point, both tapes are stopped and the original tape is rewound. The second tape is allowed to move forward about ten seconds in order to make a blank space between the first and second parts of the test. Next the original text is dubbed onto the test tape segment by segment. The segments should already be marked off in the transcription of the text. As the investigator makes this test tape he follows the broad transcription of the vernacular text to be able to determine where each segment ends. At the end of each segment, the original tape is put on "pause" while the test tape is allowed to keep running in order to insert a blank space of about five seconds between segments. This process is continued until the whole text is copied onto the test tape, segment by segment.

2.1.4 Administering the tests

The first step in administering the tests is deciding which test tapes should be played in each of the villages visited on the second round of the survey. If the survey area includes more than half a dozen different dialects it becomes impossible to administer every test tape in every village. In general one should not administer more than five tapes to any one individual or group, due to fatigue of both the subjects and the investigator. The investigator, therefore, must guess which tests will give the most information at any given village. If it is absolutely necessary that a large number of tapes be tested in one village, it can be done by playing one set of tapes to some subjects and another set to others.

To determine which dialects to test at a given village, the investigator must rely on the data which have already been collected from the area either in the planning stage or in the collecting trip. If, according to information already available, it is already apparent that the similarity between two dialects is very high, then in general there is no need to test their intelligibility. By the same token, if similarity is known to be extremely low, there generally will not be a need to test intelligibility. Also one can rely on opinions that have been collected during the first round -- the opinions of people in the villages as to what languages they can or cannot understand. The purpose of the intelligibility testing at this point is to fill in the gaps in the information, to concentrate on cases where the investigator is not sure from other evidence whether he can expect understanding or not.

The investigator may also be guided in his choice of which test tapes to administer by the characteristics of the dialect or village from which the tapes have come. Where the goal of the survey is to determine centers of communication for use in literature programs, then the investigator may want to concentrate testing efforts on the villages or dialects which might best serve as centers. This notion of centrality is based not only on linguistic and intelligibility relations but also on geography, accessibility, population, economy, politics, and the facilities (such as stores, schools, churches, clinics, airstrips) that are available in a place (see Sections 6.1.2, 6.1.4, 6.3.2; also J. Sanders 1977).

The intelligibility tests can be administered to groups of people or to individuals. Group testing can be used when the investigator can assume a homogeneity across the population as to multilingual experience; a sampling of individuals is tested when he cannot. The assumption of homogeneity or heterogeneity can be based on results at other villages in the survey and on the opinions of local people. (The topic of group testing versus individual testing is discussed in more detail in Section 2.3.2.) When individuals are tested, they should be isolated (which can be done with earphones) so that other potential subjects will not be disqualified by hearing the test and the answers. The investigator should screen the subjects to ensure that they are native speakers of the dialect, as was done for the storytellers (Section 2.1.2). The screening questions will also reveal if a subject has had a degree of contact with some other dialects which is beyond the ordinary.

When a whole group is tested at once, it is rather awkward to go around the whole group and screen the subjects first. In this case the screening can be done as the testing progresses. In group testing, a spokesman for the group will generally emerge. When questions or translations are asked of the group, the group is free to discuss and come up with an answer which the spokesman will pass on to the investigator. If it becomes clear that the spokesman or another individual is dominating a particular test, screening questions should be asked to determine if that person has had close contact with the village being tested for. If so, different individuals from the group should be asked directly for their responses to remaining segments in the test. This allows the investigator to get a sample of the understanding of the whole group.

The first tape played to any group is the test tape made of their own dialect, which is called the hometown test. This test gives the listeners the practice of taking the test without the added obstacle of dialect differences to overcome. During this hometown test, not only do the listeners have the chance to practice the test format, but also the investigator has the chance to evaluate the subjects as to their suitability for testing. It is during this hometown test that the investigator may discover deficiences in the abilities of the group or an individual subject in translating into the common language. Thus the hometown test acts not only as a practice test for new subjects, but also for a control on their bilingual abilities in the common language.

When administering a test, the first part of it, the one or two minutes of continuous text, is played without interruption. Here the listeners are given the opportunity of hearing the new dialect. The investigator may choose to withhold the identity of the dialect and see if the listeners can identify it after hearing this first section. When this first part of the test comes to an end, the investigator stops the tape and explains that now the entire story will be played from the beginning one segment at a time. At the end of each segment the investigator stops the tape during the pause and the individual subject or someone from the group is asked to translate that much of the story into the common language. If he hesitates the investigator may ask leading questions to get him started. If an important point has been omitted from the translation the investigator may ask specific questions to find out if the point was actually not understood or if it was just overlooked in the subject's translation. If none of the subjects can translate or answer any questions, ask if they understood anything, if there were any words or phrases they recognized.

The responses to each individual segment of the text should be recorded in a booklet. A convenient way to do this is to estimate the fraction of the segment which was understood, that is, record a one if all of it was understood, a zero if none of it was understood, one-half if half was understood, and so on. If only a word or a phrase was understood, that word or phrase may be written down. At the end of the test the responses are reviewed and the listeners' understanding of the test tape is summarized as being one of the four levels of intelligibility in the scale below. If a group is tested, then the understanding of the population (assumed to be homogeneous) is summarized as being of a single level. However, when it is found that an individual is dominating the answers, and then a sampling of the group is obtained to counteract, it may be reported that a few with extra experience understand at one level while the majority understand at another. If individuals are tested, then the understanding of the population is reported as the distribution of the levels of understanding among individuals. The four levels of intelligibility are as follows:

3 = full intelligibility - The listeners understood everything. At most they missed a few details of the story. In some cases a group may have difficulty responding to the first few sections, but after that they adjust to the new dialect and translate all remaining segments fully and correctly. This should be scored as full intelligibility.

2 = partial intelligibility - The listeners understood the main points of the story but missed many details. This level of understanding is characterized by incomplete understanding of segments throughout the story. The listeners understood enough, though, that they would need only to ask a few questions of the speaker to fill in the missing details. This is a level of potential full intelligibility.

1 = sporadic recognition - The listeners understood only isolated words and phrases, perhaps even occasional sentences. However, they did not know what was happening in the story.

0 = no understanding - The listeners understood nothing. Perhaps they recognized a common word like 'man' or 'house', or an important cultural item like 'betel nut'; however, there was no consistent recognition of isolated words or phrases.

Note that only the relative ordering between the levels is defined, not the relative distance between them. Thus, level 3 represents more understanding than level 2, and 2 more than 1. However, the distance between 1 and 2 is probably greater than that between 2 and 3.

2.1.5 A summary of time requirements for the method

The method requires two hours for the preparation of each test tape. When tests are administered to a group, it takes only one hour to conduct the tests in a particular village. Thus the method requires three hours per village in the survey. This compares favorably to the amount of time needed for a conventional lexicostatistic survey. However, if tests are administered to a sample of individuals, then the testing phase will take considerably longer as is detailed below.

The preparation of the test tapes for any dialect consists of the following four steps: (1) elicit the text, (2) roughly transcribe the text, (3) decide how to divide the text into segments, (4) prepare the test tape. The elicitation of the text generally takes one hour with an informant. This includes the time required to explain what is wanted, play it back for the informant, get an interlinear translation of the text, and also play that back.

The remaining three steps generally require another hour. The following time figures are based on records kept on the preparation of eleven test tapes for the dialect survey of Santa Cruz Island (Simons 1977a). Transcribing one minute of text took from five and a half to eight minutes, with an average of six and three-quarter minutes. This time includes making the rough transcription of the vernacular texts from the interlinear translation tape and then an exact transcription of the translation for each portion. Thus, to transcribe the ideal text of three minutes' length took an average of 20 minutes. After the transcription was finished, it took about ten minutes to read it over and decide where to make the breaks between segments and what leading questions could be used to prompt subjects when their response was not immediate. It took another 10 minutes to set up the two tape recorders and dub the test tape. Finally it took about 15 minutes to type up the transcription of the translation of the text with gaps in that transcription corresponding to the breaks in the test tape, and with leading questions typed into the gaps. This is a total of 55 minutes. An advantage of the method is that all of this test preparation is done without the aid of informants. Therefore, it need not be done at the test site but can be done at another place where the investigator may have set up a camp.

When the tests are administered during the second trip to the dialects, the hometown test and the four or five other test tapes can be administered to a group in one hour. If tests are administered to individuals, the process will go faster without group discussion time. About 45 minutes are required for an individual subject. That comes to three hours for four subjects, six hours for eight subjects, or seven and a half hours for ten subjects. To do a thorough job of testing intelligibility over a complete cross section of the population may require 30 to 40 subjects. Typically, testing in such depth would be done in only one or two villages out of the entire survey area in order to get a feel for the homogeneity or heterogeneity of multilingual abilities within the village populations. The in depth studies would point out the factors, if any, which explain differences in understanding (for example, sex, age, or schooling) and would give a basis for interpreting results in the rest of the suvey where only a small number of individuals were tested.

When tests are administered to a group, this method requires three hours' work by an individual investigator for each dialect. This is not much more time consuming than a conventional lexicostatistic survey. The essential difference is that the intelligibility survey requires that each village be visited twice, first to collect the test tapes, and second to administer them, whereas the lexicostatisic survey requires only one visit. However, a two pass lexicostatistic survey can give much more reliable results than a one pass survey. This is because the investigator has the opportunity to compare the word lists after they are all collected and then in a second visit to re-elicit items which appear to have been elicited incorrectly. Therefore this intelligibility approach fits very nicely with a lexicostatistic approach for the initial linguistic survey in an area. Actually, once comparative word lists are collected, analysis of phonostatistics, phonological correspondences between dialects, and lexical and phonological isoglosses can be made without collecting any additional data. When a computer is available, this added wealth of information is almost free in terms of the investigator's time (Simons 1977b describes a set of computer programs which can be used).

Provided there is not more than three hours' travel time between test points it would be possible for a single investigator to conduct the study of linguistic comparison and intelligibility at two test points in a single day. With a two man team the work becomes even easier, with one member concentrating on the linguistic side of the study and the other concentrating on the intelligibility side.

2.2 A review of intelligibility testing methods

This review of intelligibility testing methods is made in chronological order. First, in Section 2.2.1, the early studies in the 1950's are considered. Then Hans Wolff's 1959 critique of these early studies is reviewed in Section 2.2.2. This critique led to refinements in the method by a group of investigators from the Summer Institute of Linguistics in Mexico. Section 2.2.3 treats their method. Finally, Section 2.2.4 presents other recent methods.

2.2.1 The early studies

The method of intelligibility testing has its origins in a 1951 article by Carl F. Voegelin and Zellig Harris. They proposed intelligibility testing as a means of measuring dialect differences, in hopes that it could help define the border between dialect and language. Their main interest was in classifying languages rather than in communication itself. They discussed four methods which could be used to distinguish language from dialect: (1) ask the informant, (2) count samenesses, (3) structural status, and (4) test the informant. It is their "test the informant" method which has developed into the intelligibility testing techniques discussed here. Basically, their method was this: make a tape recording in dialect A and see how well speakers in dialect B can understand it. Voegelin and Harris suggested measuring understanding by noting the accuracy with which speakers of dialect B could translate the text.

Hickerson, Turner, and Hickerson (1952) were the first to use the Voegelin and Harris method of testing the informant in a field study. They refined the sketchy outline of the method given in the original paper to determine relationships among seven Iroquois languages of North America. A second intelligibility survey was conducted soon afterwards by Pierce (1952) among Algonquian languages of North America. Later Biggs (1957) conducted a similar survey among the Yuman languages of North America.

All three of these surveys used basically the same method. The investigators obtained a translation of the original text and then scored the subject's translation of that text to arrive at a percentage of items which were correctly understood and translated. In the first study, the investigators took down an exact translation of the text from its teller, and scored section by section translations of the text by subjects as incorrect, one-third, two-thirds, or fully correct.

Pierce (1952) used what he called a "standard grading translation". Rather than trying to obtain an exact morpheme by morpheme transcription and translation, he obtained a running translation into English from the person who told the story and from two other speakers of the same community. He then compared these three translations to construct the standard grading translation which listed the main semantic units in each sentence of the text. A subject's translation of the text was scored as correct, incorrect, or half correct for each unit. Pierce also recognized the importance of the hometown test as a measure of a subject's abilities and was the first to suggest that it could be used in adjusting raw intelligibility scores to control for differing subjects' abilities. Biggs (1957) was the first to administer tests to groups of subjects.

In these three studies, as well as in the original proposal of Voegelin and Harris, the main emphasis or perspective was language classification. They were not interested in the intelligibility scores as a measure of communication as much as they were interested in using intelligibility to measure "dialect distance" (Pierce 1952, Biggs 1957) -- the degree of relatedness between speech groups. Because of this they took the asymmetry out of intelligibility test results by computing what they called a percentage of "mutual" intelligibility, which averaged the amount of information flow in both directions between a pair of dialects.

2.2.2 Wolff's critique

In 1959, Hans Wolff wrote a criticism in which he questioned the validity of using intelligibility testing to measure "dialect distance". In his paper he makes the following criticisms of the method (this list follows Yamagiwa 1967:14-15):

(1) The method seems to measure primarily the subject's ability to translate. While ability to translate obviously presupposes some type of intelligibilty, the reverse is not necessarily true.

(2) The translation is made into a third language, thus introducing an additional uncontrollable factor.

(3) The subject may dislike the notion of having to produce a translation.

(4) The subject's reaction to hearing speech from a lifeless box rather than in a normal sociolinguistic situation constitutes another uncontrollable variable.

(5) The subject's psychocultural reaction to a different form of speech and possibly to the people who customarily speak it may enter into the testing.

(6) Dialect distance can be tested effectively only if the non-native dialects have not been learned.

(7) The test does not permit us to distinguish between intelligibility due to linguistic proximity alone and that which is due to some kind of learning process.

(8) The test yields little useful information when we are faced with the baffling phenomenon of nonreciprocal intelligibility.

Wolff went on to discuss the cultural factors involved in communication between different dialects, illustrating with four examples from Nigeria. He concludes that although linguistic proximity may play a limiting or boosting role in communication, the decisive factors are cultural. Thus intelligibility tests could not be a valid means of measuring linguistic proximity.

While all of the points made by Wolff are basically correct, he made a few oversights which render many of his criticisms vacuous. In the first four points he spoke of uncontrollable factors which affect the results of the intelligibility tests: the subject's ability to translate, the subject's proficiency in a third language, the subject's dislike for having to produce a translation, and the subject's reaction to hearing speech from a lifeless box. Here Wolff's use of the word "uncontrollable" is incorrect. What he means is "unmeasurable". It is true that the subject's translation ability, his bilingual ability, and his attitude toward the test situation cannot be easily measured. However, these measures can be controlled for and an attempt to do so was made in the early intelligibility studies.

The hometown test (the test on the subject's own dialect) was used in the early intelligibility tests for this purpose. Pierce (1952:206-7) goes to some length to explain how the hometown test can serve as an experimental control for these unmeasurable factors. Presumably, an informant should score 100% understanding of his own dialect. Any difference between the observed score and 100% can be attributed to the factors above: a lack of ability in the translation language, a lack of skill in translation, or a reaction against the test situation. Pierce suggests we can assume that these same kinds of deficiencies which affected the subject's translation of his own dialect will also affect his translation of the other dialects. Pierce goes on to say that if all of a subject's scores are divided by his score on his own dialect, the unmeasurable factors cancel each other out. As a result the score on his own dialect will be raised to 100% and all other scores will be raised by a proportional amount. All such scores between different informants are comparable because the effects of differing levels of subject ability have been compensated for. As Pierce shows, we may not be able to measure exactly the effect of translation skill or a third language skill in the test results, but the fact that we divide the one score by the other cancels out their effects and what remains is the measure of intelligibilty. Thus Wolff failed to recognize the significance of the hometown test as an experimental control in the intelligiblity test design.

The eighth point above, that the test yields little useful information when we are faced with the baffling phenomenon of nonreciprocal intelligibility, is not a criticism of intelligibility testing itself, but rather is a criticism of the way the early studies interpreted the results of the test. Pierce (1952) and Biggs (1957) disregarded the asymmetry in intelligibility relations. Since "dialect distance", the relation which they were trying to measure, is symmetrical, they chose to compute a "percentage of mutual intelligibility" as a measure of dialect distance. This percentage was the average of the score in each direction between two dialects. The fault here was not in their method of measuring intelligibility but in their assumption that it should be "mutual" when in fact, it is not.

The remaining three criticisms are again a criticism not against the method of testing intelligibility, but against the way in which the original investigators interpreted and applied their results. Wolff was arguing that intelligibility scores not only tell us something about the linguistic distance between two dialects but they also tell us something about the social relations between the dialects. This relation could be manifest in attitudes which would result in a negative kind of reaction against the test tape (point number 5), or in favorable kinds of relations that could result in the learning of different dialects (points 6 and 7).

With four examples from Nigerian languages Wolff goes on to show that intelligibility measures both linguistic relations between dialects and social relations. Wolff left off his argument at that point. It follows, however, that if the investigator can demonstrate that the social relations between dialects are absolutely nil, then the measure of intelligiblity can be viewed as reflecting only linguistic relations. In such a case, intelligibility scores may have value as offering a composite measure of phonological, lexical, grammatical, and semantic relations between dialects. It was such an understanding that motivated Biggs (1957:59) to screen his subjects and discount any subject who had had extensive prior contact with the language being tested.

An analysis of Wolff's criticism of intelligibility testing as it was practiced in the 50's indicates that he actually made no legitimate criticisms against the method that was being employed to measure intelligibility, only against the way the results were interpreted. The real value of his paper is in demonstrating with some very good examples that intelligibility measures not only linguistic relations but also social ones. Thus we must be extremely cautious in interpreting intelligibility scores as a measure of dialect difference.

2.2.3 Casad's method

In the early 60's, John Crawford began adapting the methods of intelligibility testing (Casad 1974:58 ff). He agreed with Wolff's criticism of the way the early intelligibility tests had been administered and interpreted. However, he reasoned that if the actual testing technique were improved, intelligibility scores could be used as quantitative measures of the amount of information transfer between dialects. His interest was not in dialect distance but rather in how widely a dialect could be used in vernacular literature and education programs. Papers by Bradley (1968) and Kirk (1970) are initial reports on Crawford's refined technique and its application in a number of projects by the Summer Institute of Linguistics in Mexico. They took his ideas and continued to refine the techniques of intelligibility testing while conducting field surveys of various language groups.

Three works which have recently come out of that project give extensive coverage of dialect intelligibility testing. Casad (1974) has written a thorough manual on how to conduct a survey and how to interpret the results. In addition he gives historical and critical reviews of the method and discusses alternative techniques. His book is an invaluable source on the topic of dialect intelligibility testing. Stoltzfus (1974) treats the problem of designating certain dialects as centers for indigenous literature programs and then supports the discussion with analyses of six dialect surveys conducted in Mexico. Grimes (1974) concentrates on the methods used to analyze the survey data and convert them to decisions on dialect groupings and centers.

In this approach, Wolff's major criticism, that intelligibility scores are not valid measures of dialect distance, is bypassed by viewing the scores strictly as measures of information transfer, not dialect distance. Here the investigators are interested in determining the extendability of vernacular literature produced in any given dialect. The social factors which affect intelligibility, such as negative feelings that limit communication or good relations that boost communication, are also likely to limit or boost the extendability of literature in the same way. Thus intelligibility taken as a composite measure of both linguistic and social relations, measures exactly what they were looking for.

These investigators also tried to combat the other aspects of the original test design which Wolff criticized. Wolff criticized the method because it tested a subject's ability to translate as much as it did his understanding. Thus the method was changed so that subjects are not required to translate a passage, but to answer specific questions about the content of a passage. As a convention, a text is divided into ten segments and a question is asked for each one. Moreover, Wolff complained that the subjects were required to make a response in a third language. In this refined technique the questions and the subject's answer to these questions are given in his own vernacular. Again, Wolff felt that the subject's reaction against the test situation itself and against the methods and equipment of the investigators could introduce an uncontrollable variable into the test results. In this refined technique the first tape which any subject listens to is an introductory tape in his own dialect. This tape first introduces the investigators and explains their purpose, then it explains how the testing will be done and gives a short sample test in which questions are asked and the correct responses are given for an example. This introductory tape is meant to relax the subject and familiarize him with the investigators, their techniques, and their equipment.

Casad summarizes the steps in preparing the test tapes for each dialect as follows (1974:100):

> The survey team must complete the following series of steps at each test point: (1) elicit and transcribe an adequate text, (2) formulate a set of questions from the translation of that text, (3) translate the sets of questions for all the test tapes into the local dialect, (4) prepare an introduction tape, (5) submit the translations of the questions to a pre-test panel of speakers of that dialect in order to detect and correct translation errors, (6) make a dubbed copy of a hometown text for constructing the hometown test tape, and (7) record the translated questions and the introduction tape.... This preparation entails a day's work.

To administer the tests requires another day. The tests are administered to individual subjects and as a convention, Casad suggests that ten subjects be tested. About 45 minutes are required to administer a set of test tapes to a single subject (Casad 1974:24). That amounts to seven and a half hours for ten subjects. Furthermore, the method requires a survey team of two members (1974:3). The total requirement for each test point is then two investigators for two days, or four man-days.

2.2.4 Other recent methods

In this section, four other recent methods are considered. In each case these investigators studied intelligibility to learn about communication between speech communities and not to estimate dialect distance. Thus they sidestepped the brunt of Wolff's criticism. In the first three methods the investigators used written tests to increase the efficiency of data collection. In the second, third, and fourth methods the investigators used translated texts to control for variation in the difficulty and subject matter of the test materials.

Yamagiwa (1967) studied intelligibility among Japanese dialects by administering tests to sixty-five university students and graduates. Because of the academic sophistication of his subjects, he was able to have the subjects make written translations of the texts they heard. The students heard short portions of speech twelve to twenty seconds long recorded in ten different dialects. They heard three repetitions of each dialect sample and were asked to write out a translation of each. These translations were compared to a standard translation to score the amount of understanding. The advantage of this written translation method is twofold: tests can be administered in a classroom kind of setting so that a large number of subjects can be tested at one time, and each subject transcribes his own translation so that the investigator is not left with recordings that he must later transcribe and score. The effect is that the investigator can collect many times more information in much less time than is possible by the methods in Sections 2.1, 2.2.1, and 2.2.3. Of course such a method is limited to a very restricted kind of subject.

The next two studies come out of the "Survey of Language Use and Language Teaching in Eastern Africa" project. They are the study of intelligibility among the Sidamo languages of Ethiopia by Marvin Bender and Robert Cooper (1971) and tests conducted with speakers of two Bantu languages of Uganda by Peter Ladefoged (1968, Ladefoged and others 1972). Bender and Cooper studied intelligibility among six languages. Six separate stories about everyday topics were translated into each of the six languages, giving thirty-six passages. These were then spliced into six test tapes, with one story per language on each tape. The order of the languages was different on each tape. The tests were administered to sixth grade school children in the classroom. On the test tapes each story (averaging 175 words) was followed by three questions with four multiple choice responses each. The questions and their alternative responses were printed in test booklets in Amharic. The students heard each story with its questions and responses two times before marking the response in the test booklet. Before the actual testing began there were three practice exercises. Bender and Cooper report that it took 45 minutes to administer a set of six tests, including practice. That is, they were able to test a whole classroom full of literate subjects in the same amount of time that one subject can be tested by the methods reviewed in preceding sections.

The use of translated stories has the disadvantage that the investigator cannot insure the naturalness of the texts. On the other hand, it has the advantage that one can experimentally control for the differing difficulties of the texts. They translated six stories into six languages, or thirty-six passages. These were arranged in six test tapes including one passage for each language. Students to take the tests were divided into six groups and each group heard a different test tape. The result was that each group heard a different story from the same language. The sum of responses for all six groups on a particular language is a sum over all six stories. The total responses on each of the six languages are a sum over the same six stories, and therefore the intelligibility totals are based on identical texts and questions. This is not true of the other methods considered so far. This particular kind of experimental design is called a Latin Square design. Coupled with the statistical method of analysis of variance, it can be used to test hypotheses concerning the relative effects which different groups of subjects, different stories, different languages, and the ordering of the tests on each tape have in explaining the observed differences in intelligibility.

Ladefoged used basically the same method as Bender and Cooper with translated stories, Latin Square design, and written tests among school children. His tests were simpler in that the stories were shorter and the subjects were required to answer only one multiple choice question on what each story was about. Each question had three possible responses.

Gillian Sankoff (1968:151-5) used translated texts to test intelligibility among the Buang of Papua New Guinea using an oral method. She made tests for six languages. First, she composed six short stories having to do with daily life in the village. The English text of each was about 100 words long. Then each story was translated and recorded on tape in each of the six languages, resulting in thirty-six taped stories. Each subject then heard a different combination of texts in the different languages, though the order of the languages was kept constant. In administering the tests, an individual subject listened to a story once. Then the investigator personally (rather than on the test tape) asked three questions about it in the vernacular. The questions were phrased so as to have brief answers. An answer was scored 2 for completely correct, 1 for partially correct, and 0 for wrong. The result on a test ranged from 0 to 6. In the three dialects she tested 16, 20, and 48 subjects. A drawback of the method was that subjects tended to forget items which they actually understood, as evidenced by the fact that subjects averaged 70% understanding of their hometown dialect test.

A second part of the test was designed to measure comprehension of vocabulary items in the texts. Ten content words were selected from each story for the test of vocabulary items. After the subject had listened to the story and answered the questions the text was played again, stopping the tape at the ten selected words. The subject was then asked to translate the word into his own dialect. A response was scored as correct or incorrect, thus 10 points were possible for the test.

2.3 A taxonomy and evaluation of intelligibility testing methods

The methods of intelligibility testing which have thus far been presented are now classified according to six dichotomies. The alternate approaches within the dichotomies are evaluated in terms of optimality and relation to the investigator's goals. As a conclusion the relation between the abilities of the potential subjects and the methods of testing is considered. The results show that no one method of testing intelligibility is inherently better than another. Rather, the goals of the investigator and the capabilities of the subjects work together to define a method which is best for a situation. It is hoped that the following discussion will serve as a guide to those who must plan a dialect intelligibility survey.

2.3.1 A taxonomy of intelligibility testing methods

Methods of intelligibility testing can be classified according to the following six dichotomies:

(1) Language of response = (Vernacular, Common)
(2) Mode of response = (Oral, Written)
(3) Format of test = (Question, Translation)
(4) Scoring method = (Quantitative, Qualitative)
(5) Source of text = (Elicited, Translated)
(6) Sampling method = (Groups, Individuals)

That is, (1) subjects may be asked to respond in their vernacular language or in some language such as the national language or trade language which is common to them and the investigator, (2) subjects may be asked to speak their responses or write them, (3) the test may be formatted so that subjects are asked to respond by answering questions about the text or by translating it, (4) understanding may be scored quantitatively (as a percentage, for instance) or qualitatively (as being adequate or not adequate, for instance), (5) the texts for tests may be elicited narrative or may be translations of pre-written texts, and (6) the subjects may be sampled as a group or as individuals. Figure 2.1 sets out a table in which the methods discussed thus far in the chapter are classified as to their values for each of the six dichotomies.

2.3.2 An evaluation of intelligibility testing methods

There are many possible ways to test intelligibility. The methods classified in Figure 2.1 are ways that have been used; other combinations of the six variables could be proposed. In this section I argue that no one method is inherently better than another. This observation is borne out in Section 5.5 where it is shown that the many different methods give essentially the same result. The choice between methods is therefore based on restrictions caused by the abilities of the subjects (see Section 2.3.3) and by the investigator's goals. Where choices still remain, the decision is based on a criterion of optimality. I define an optimal method as one which allows the investigator to gather the greatest amount of information possible with the least amount of effort possible. Each of the six dichotomies is now considered in turn.

(1) Language of response - The subjects can respond in their vernacular language or in a common language, such as the national language or a trade language. In a question approach, the same language will also be used to formulate questions. Where a common language can be used (that is, where the subjects are adequately bilingual), the common language approach is optimal. This is because it requires the least amount of effort to prepare test tapes. In a vernacular approach, such as Casad's (Section 2.2.3), it is necessary to construct a new test tape for each village where a text will be tested. The text remains the same but the questions that go with it must be translated into the local dialect and dubbed in to create a new test tape. Otherwise the difficulty of understanding the questions themselves compounds the difficulty of understanding the text. With the common language approach, however, the same test tape is used in every village where a particular dialect is tested. This saves time as well as increasing consistency since subjects in different villages hear the same test tape instead of different versions of it.

One of the problems associated with selecting a common language approach is insuring that the subjects are adequately bilingual in the common language. When a common language approach is used, the hometown test serves as a control for bilingual abilities. If a subject's score on the hometown test does not near 100%, then it may indicate that he is not sufficiently bilingual to take the test. However, if a subject does score nearly 100%, then his bilingual abilities are not at issue. Thus the hometown test serves to validate the assumption that the subjects are sufficiently bilingual to be tested in the common language. However, if it turns out that many potential subjects are disqualified from further testing because they cannot respond adequately on the hometown test, then the use of the common language may bias the results. If the investigator still wants to use a common language approach, then it may be necessary to do an in depth study in one or two villages with a vernacular method to see if a common language approach for the entire survey would be valid. For instance, Gillian Sankoff (1968:168-9) used a vernacular approach to test 48 individuals from Mambump village, Papua New Guinea, on their understanding of related dialects and on three languages of wider communication (including New Guinea Pidgin). She found that although women scored significantly lower than men on the Pidgin test, there was no significant difference in their scores on related dialects. Thus I would suggest these results indicate that in this area one could use a common language approach with Pidgin which tested

Figure 2.1 Methods of testing.intelligibility

	Language of response	Mode of response	Format of test	Scoring method	Source of text	Sampling method
Hickerson and others 1952, Pierce 1952	Common	Oral	Translation	Quantitative	Elicited	Individuals
Biggs 1957	Common	Oral	Translation	Quantitative	Elicited	Groups
Casad 1974	Vernacular	Oral	Question	Quantitative	Elicited	Individuals
Simons 1977a	Common	Oral	Translation	Qualitative	Elicited	Groups
Sankoff 1968	Vernacular	Oral	Question	Quantitative	Translated	Individuals
Yamagiwa 1967	Common	Written	Translation	Quantitative	Translated	Individuals
Bender and Cooper 1971, Ladefoged 1968	Common	Written	Question	Quantitative	Translated	Individuals

only men, and not bias the results by excluding women from the sample.

Another way in which a common language approach is optimal is that it does not require that the investigator be familiar with the vernacular language. Casad suggests that for his method one of the investigators should speak at least one of the dialects in the language area under study (1974:3). A method with that requirement means that intelligibility cannot be tested in a new linguistic area until someone has spent a number of months, perhaps a few years, learning a language. Joseph Grimes (personal communication) feels that until there is such an investigator, intelligibility is too fine grained a phenomenon to measure. I do not think so. There should be methods of testing intelligibility (some are suggested in Section 2.3.4) which allow a team of survey technicians to go into any new area and survey the intelligibility situation (as well as the linguistic and social relations, Chapters 5 and 6) so that wise decisions about language planning can be made before personnel are actually assigned to in depth study of languages in the area. The in depth study may later suggest some changes in strategy, but they will not be as drastic as they would have been had initial decisions been based on linguistic relations alone. A common language approach, where it is applicable, means that the intelligibility situation can be measured immediately without having to wait months or years for the investigator to gain proficiency in the local vernacular. Where the common language approach cannot reach all of the population but yet a sizeable portion of it, then it is still valid to use it with the understanding that it gives better results than no survey would, and that a vernacular approach will be used at a later time to refine the analysis of the dialect situation.

(2) Mode of response - The subjects may respond by speaking or by writing. Where it can be used, the written approach is certainly optimal. Bender and Cooper (1971) could simultaneously test all of the students in a classroom on six written tests in 45 minutes (Section 2.2.4). Casad (Section 2.2.3) and myself (Section 2.1.5) could administer such a battery of oral tests to only one indivudual in the same amount of time. Where the level of writing skill is high enough not to form a barrier of its own, written tests yield a much higher return for a given amount of effort than oral tests.

(3) Format of test - The test may be formatted so that subjects can respond by answering questions about the content of the text or by translating it. In this case, we cannot really claim one method is optimal over another; this depends on the other methods being used alongside it and the investigator's goals. A translation approach is optimal as compared to a question approach in which questions are dubbed into the test tape because it is simpler to prepare the test tapes for the translation test. On the other hand, it is much much easier to score a question test quantitatively than it is to score a translation test. In the case of the method I suggest in Section 2.1, the translation appproach turns out to be optimal since a qualitative method of scoring is used. There is a tradeoff here between the two aspects of optimality: information and time. A translation approach measures understanding of every item in the text, not just selected points of content, and thus yields more information. A question approach, on the other hand, yields its information with much less effort since it does not require item by item comparison of translations.

The choice between translation and question approaches is partly one of sampling. In a question approach, the investigator is sampling from the text. He is concentrating on a few points of content and trying to generalize to the whole text. The problem is essentially, "What is the likelihood that the subject's responses to the selected questions are a good indication of his understanding of the whole text?" The accuracy with which the questions reflect the whole text is involved here. The questions may happen to hit only items which are similar between the dialects, or only items which are dissimilar. The more questions that are asked, the less likely the sample will be biased.

The phrasing of the questions will also affect the responses. That is, a subject may have understood something but the question does not bring out that fact. In a study conducted by my wife and me among dialects of the Biliau language in Papua New Guinea (L. Simons 1977:250), we found that the most often missed questions were "why" questions. The second most often missed questions were "how" questions. The third most often missed were questions where the answer was rather far from the end of the segment of text. We thus found that questions can be simple or difficult depending on how they are phrased and where their answers are found with respect to the gap in the test tape. These phenomena are independent of the subjects' understanding of the text. However, they will surface in the responses given by the hometown dialect. These factors can be controlled for by adjusting all of a test's scores on the basis of its hometown score (Section 5.2.4).

When a translation approach is used, understanding of the whole text is tested. This avoids the sampling problem of how well understanding of the items questioned measures understanding of the whole text. However, it still does not avoid one serious sampling problem which affects all methods of intelligibility testing. This can be summed up in the question, "What is the likelihood that a subject's understanding of this text is a good measure of his understanding of the whole language?"

(4) Scoring method - A subject's understanding of the text can be scored quantitatively or qualitatively. In a quantitative method, the number of items correctly translated or the number of questions correctly answered is added up and the resulting number is the score. These scores are generally converted to percentages. In a qualitative method the investigator does not count the understanding; rather, he judges it along some discrete scale of levels, for instance, adequate or not adequate; full intelligibility, partial intelligibility, or no intelligibility. Again we cannot pronounce one method optimal in all cases. With translation approaches the qualitative approach is optimal in the sense that it requires less effort; however, a quantitative approach opens up a broad range of statistical methods that can be used in the analysis of results. With question approaches, the questions may not provide a large enough sample of the text to allow a qualitative judgment. For instance, the written methods of Bender and Cooper (1971) and Ladefoged (1968) presented in Section 2.2.4 used only three questions and one question respectively. It would be impossible to base qualitative judgments on such small samples.

Of all the methods classified in Figure 2.1, the method I suggest in Section 2.1 is the only one which uses a qualitative method of scoring. Since such a method of scoring has not appeared widely in the literature it would be good to discuss it here. Qualitative scoring is of advantage because the scores have an interpretable meaning in the real world. Also, qualitative scoring avoids one of the problems of quantitative scoring, overprecision.

When the investigator scores intelligibility qualitatively, he knows what the scores mean and how they should be interpreted in applying intelligibility test results. For instance, if intelligibility is scored on a simple dichotomy as adequate or not adequate, the investigator knows which intelligibility relations are

31

adequate for establishing dialect groupings and which are not (Chapter 3). When intelligibility is scored quantitatively, however, the step of interpreting scores for applying results still lies ahead. What does it mean if subjects score 70% intelligibility? In the early methods (Section 2.2.1) it means that a subject was able to correctly translate 70% of the text. In Casad's method (Section 2.2.3) it means that on average the subjects answered 7 out of 10 questions correctly. In Ladefoged's method (Section 2.2.4) it means that 70% of the subjects answered the one question correctly. To go from these measurements of 70% to what they mean in terms of levels of communication adequacy for a vernacular language program (Section 3.1) still requires a step of subjective interpretation. When scoring is done qualitatively this subjective interpretation occurs at the test site as the test is administered rather than weeks or months later when the results are analyzed.

A potential pitfall of quantitative scores for the unwary investigator is that they are overprecise. That is, the percentage scales which have customarily been used appear to discriminate 100 degrees of intelligibility. In actual fact they do not. Statistical tests of significance show that even 10% differences in measured intelligibility need not be significantly different. In an appendix to Casad's manual (1974:167-173) the standard deviations as well as the means for inttelligibility scores from the Mazatec survey are reported. A one-tailed *t* test shows that Tenango's hometown score of 95% (6.71% standard deviation) is not significantly greater at a 95% confidence level than the score of 87% (12.69% standard deviation) which another test point, TE, scored on Tenango. TE's score of 46% on the Jalapa test is not significantly greater at a 95% confidence level than MZ's score of 35%. On the other hand, TE's score of 90% on San Jeronimo is significantly greater at a 95% confidence level than HU's score of 76%.

These tests of significance are actually testing the hypothesis that one group's score on a test is greater than another group's score on the same test. They are not testing the hypothesis that one group's intelligibility of a dialect is greater than another's. To test the significance of the difference between a group's score on one test and its score on a second test may not even be possible since the tests are different. We would have to know how the two tests compared with respect to a language sampling distribution. The significance tests made above take into account only the variation in subject sampling. To make inferences about intelligibility, not just test scores, we would have to take into account variation in language sampling as well. Unfortunately we have no way of measuring this.

Casad, in an appendix (1974:173), suggests that we might do better to state results in terms of range estimates rather than point estimates. Grimes, in a footnote (1974:262), suggests that decisions concerning intelligibility test results "should ultimately be based on tests of the significance of the differences between two ranges rather than on the simple greater-than, less-than relationship between two numbers." Both are correct; unfortunately, these suggestions have yet to be implemented in field studies.

The use of qualitative scoring techniques offers a way out. On a qualitative scoring scale, all the levels of intelligibility are significantly different since there are so few levels, generally fewer than five. When there are no significant differences in the distribution of intelligibility within the whole population, then the qualitative level of intelligibility is reported, and there is no range or distribution to report. If there are significant differences in the distribution, then that distribution is reported; for instance, half of the population understands at a level of partial intelligibility and the other half understands at full intelligibility. Such a statement is easier to interpret than one with a percentage and standard deviation; for instance, the average degree of understanding in the population is 75% with a standard deviation of 20%. Of course, with such a reduced number of levels on a qualitative scale the problem of borderline cases can arise; that is, cases which are simultaneously not significantly different from two adjacent levels. In these cases, we should probably not treat them as different from either level, but as occurring in both, for analysis purposes (Chapter 3).

Note that using a qualitative scoring technique does not give the same result as using quantitative scoring and then reducing the results down to a four or five point scale. The latter method depends on finding discrete breaks in the distribution of test scores or simply rounds scores without regard to breaks; the former method relies on everything the investigator knows about a situation (including where the particular test might occur in a language sampling distribution) to make the judgment. The qualitative method would probably overlap in the border regions if compared to strictly quantitative results.

Although qualitative scoring scales have an advantage in terms of interpretability, quantitative scales are more advantageous in another aspect: their amenability to statistical methods for modeling purposes. I discovered this advantage of quantitative scores while working on Chapters 5 and 6. In Section 6.3, intelligibility is measured on a four point qualitative scale and the functions which predict intelligibility are step functions. Statistical methods like correlation and regression are not appropriate for the data. In Chapter 5, percentage measurements of intelligibility are used and the scope of statistical methods available for the analysis is very broad.

Both quantitative and qualitative scores have their advantages and thus we might do best to record both. After an investigator has finished gathering quantitative results on a test, he could make a qualitative judgment concerning the degree of understanding. These judgments would be used in the analysis stage to give meaning to the percent scores for the sake of interpretation.

The results in Chapter 5 illustrate something of the paradox surrounding the use of quantitative versus qualitative scores. In Section 5.4 the relations underlying percentage of intelligibility and percentage of lexical similarity are very nearly the same in eight out of ten field studies. This gives credence to the original use of percentage measurements. However, in Section 5.6 (Figure 5.8), when these data are pooled, the standard error of estimate for predictions of per cent intelligibility is plus or minus 13%. This amounts to plus or minus 26% for a 95% confidence band (see Section 4.4). This wide variation, in turn, suggests the desirability of a discrete qualitative scale.

(5) Source of text - The intelligibility tests may be based on texts which are elicited free narrative or on translations of predetermined texts. In terms of time and effort, the elicitation method is optimal. It is easier and faster to elicit a free narrative than it is to elicit a correct translation. Another advantage of a free narrative is that the investigator can be reasonably sure that the syntax, vocabulary, and semantics are natural. With a translated text he cannot. However, the use of translated texts does have the advantage discussed already in Section 2.2.4. With translated texts the investigator can use a Latin Square experimental design to control for variations in language sampling. Although he still cannot ensure that intelligibility on the texts adequately measures intelligibility on the whole language, he can ensure that all the measures between different dialects test the same sampling of language.

(6) Sampling method - The subjects can be sampled as individuals or as a group. To administer the tests once to a group is optimal in the sense that it can be done more quickly. To administer the tests to a number of individuals is optimal in that it yields more information. If written tests are used, subjects would always be sampled as individuals. In the case of oral tests, the decision is based on the goals or assumptions of the investigator. If the investigator wants to know only about the potential for understanding or if he can assume that the population is homogeneous in its multilingual abilities then a group method can be used. If he wants to know about how a particular community varies in its ability to understand, then an individual method is necessary.

Biggs (1957) was the first to use a group testing method. In such a method a number of subjects are tested collectively rather than individually. There is generally a spokesman for the group who makes the responses to the investigator. The members of the group are allowed to converse in their vernacular language before making their responses to the investigator. In using this method, Biggs argued that the scores obtained would be near the upper limit for intelligibility between the dialects. Allowing discussion between subjects allows their best responses to come through. Therefore, a group responding together is likely to score higher than the average of all the subjects responding individually. It allows the subjects to score nearer their potential. The result is perhaps more like what individuals would score if they had an hour or a day to listen instead of just two minutes. The danger is, of course, that a few individuals who have learned the other dialect will dominate the whole test while those who do not understand remain silent. If the investigator senses that this is happening, he must ask other specific individuals to respond in order to get a sampling of the group. When this is done, the investigator can actually record more than one score for the single testing situation. He can observe that the majority understands at one level, while a few understand at another higher level. Also the investigator may note the group's response to the story. Are they attentive? Do they laugh when it is humorous? When a qualitative judgment is made, the investigator need not rely exclusively on spoken responses.

Group testing may hide variablility in the population. When a group test is used to sample a population, one cannot observe much more than that some understand at one level while others are at another. More exact methods of sampling are required in order to make more precise statements about different levels of understanding throughout the population. When some individuals understand another dialect better than others, it is because not all individuals have had the same amount of contact with the other dialect. If investigations have shown no reason to suspect contact, if contact can be assumed to be uniform (for instance on the basis of preliminary tests on individuals), or if the interest of the investigator is in the upper potential, then group tests are appropriate. If the investigator wants to know precisely how the population varies in its abilities to understand the other dialect, then he must use tests on individuals. The results of such tests can be compiled into a profile of multilingual abilities in the community. Gillian Sankoff (1968:169-173, 1969:846) has done this for the Buang of Papua New Guinea. In three different plots she shows how the Mambump community's understanding of other Buang dialects and of national and regional languages varies with sex, age, and level of schooling.

When a dialect's intelligibility abilities are not homogeneous and individual testing is used to discover what the composition is, then sampling becomes an important issue. Sankoff tested 48 speakers in order to build the profile of the Mambump dialect's multilingual skills. When the goal is to see how intelligibility varies with different factors in the population, it is necessary to get a good stratified sampling with respect to those factors (Miller 1977). For instance, if differences in the understanding of men and women are to be compared then ideally equal numbers of men and women should be tested. If age differences are to be investigated, then equal numbers in each age bracket should be tested. The sample chosen should represent a cross section of the whole population.

In Casad's method, where ten subjects are tested, the size of the sample is not sufficient to make inferences about the profile of the population. It can only establish whether or not there is variability. Unfortunately, none of the intelligibility surveys on which Casad reports have taken advantage of the fact that ten subjects were tested in order to conclude something about the variability in the population. Thus far they have considered only the average of the ten scores. Casad does compute some standard deviations to illustrate a measure of variability in an appendix on statistical measures (1974:170). In the set of thirteen scores, the standard deviations for scores above 90% (including hometown scores) range from 6% to 8.5%, and for scores below 90% they range from 12% to 20%. Before any inferences can be made about how large this variation actually is, the standard deviations must be adjusted to account for the deviations in the hometown test (see Sections 2.2.2 and 5.2.4 for adjustment of means). If there is a scatter in the hometown results amounting to an 8% standard deviation, then we can assume that whatever factors caused this scatter will cause at least that much scatter in other tests. Whatever causes scatter in the hometown test is not intelligibility; all subjects should theoretically score 100% intelligibility on their own dialect with no deviations. Adjusting non-hometown standard deviations in the above Mazatec example would cut them down by about half.

Casad (1974:171-3) goes on to show how the standard deviation is used to compute the standard error of the mean and then to construct a confidence interval, or range estimate, for the mean. When the intelligibility for a population is reported as a range estimate, it is saying that the average intelligibility for the population lies between two values with a given degree of confidence. It occurs to me that this treatment of the results is actually hiding the variability which it seeks to account for. It is assuming that what we really want to know is the average intelligibility, so it accounts for variability by saying that the average lies within a range.

What the language planner needs to know is not the average nor its range, but the distribution of intelligibility. The planner may be interested in how well those at the low end of the distribution scored, or he may be more interested in the upper potential indicated at the high end. He may want to define the level of intelligibility for a population as a median (rather than a mean) which says 50% understood better and 50% understood worse, or he may want to pick some other percentage. For instance, he may think it better to characterize the population by a level which has 80% of the population understanding that well or better and 20% understanding below that level. He may be interested in the differences between sexes or he may want to concentrate on the responses of a certain age group. All of these possible applications of survey results require a method that is sensitive to the distribution of scores within a population. This area may prove to be the next frontier in the refinement of intelligibility survey methodology. Thus far the work of Sankoff (1968:164-176; 1969:846) serves as our only model.

2.3.3 Subject profile and optimal methods

The choice of which intelligibility testing method is optimal for a particular situation has a lot to do with the capabilities of the potential subjects. In this section, the relation between the subjects and the choice of an optimal method is discussed.

Potential subjects can be classified according to two dichotomous variables. First of all they may be classified according to language proficiency as monolingual or bilingual. Specifically, bilingual means fluent in a common language like a trade language or national language which the investigator also speaks, monolingual means that they share no such common language. Second, the subjects can be classified according to reading (and writing) proficiency as literate or illiterate. These dichotomies are summarized as follows:

Language proficiency = (Monolingual, Bilingual)
Reading profiency = (Illiterate, Literate)

In actual fact these are not dichotomies, but continua, and the two values given are the end points. The investigator must evaluate where the subjects as a whole fit on the continuum and decide, for instance, if they are bilingual enough to use a bilingual testing method or if they require a monolingual one. This point is considered in more detail below.

The abilities of the subject will partly dictate the method of testing used. These two aspects of subject abilities interact directly with two of the testing variables, language of response and mode of response. That is, monolingual subjects require a vernacular approach while bilingual ones could use either a vernacular or a common language approach. Furthermore, illiterate subjects require an oral approach while literate ones could use either an oral or a written one. The optimal method for each of the four possible combinations of the subject capabilities are as follows:

Subjects: Monolingual, Illiterate
    Possible methods: Vernacular, Oral
    Optimal method: Vernacular, Oral

Subjects: Monolingual, Literate
    Possible methods: Vernacular, (Oral, Written)
    Optimal method: Vernacular, Oral

Subjects: Bilingual, Illiterate
    Possible methods: (Vernacular, Common), Oral
    Optimal method: Common, Oral

Subjects: Bilingual, Literate
    Possible methods: (Vernacular, Common), (Oral, Written)
    Optimal method: Common, Written

When the subjects are monolingual and illiterate then only one method is available, a vernacular and oral method. If monolingual subjects happen to be literate, then a written approach would be possible as well. However, it would not be an optimal method in the case of a normal intelligibility survey. A vernacular written approach would require that a different set of test booklets be printed up for each dialect where tests are conducted. It would take less time to administer tests orally to a sampling of individuals, than it would take to prepare the test booklets and then test all subjects at once in a classroom sort of situation. The exception to the optimality of an oral approach for monolingual literates would be if the goal of the survey were not so much to test intelligibility between the dialects in an area as to compile a detailed profile of the multilingual abilities of a few dialects. In this case the investigation would be more like a census than a survey and the printed test booklets would pay off.

When the subjects are bilingual, then common language approaches are always optimal. The reasons have been already been discussed in the preceding section: the test materials are prepared only once and the investigator need not invest months or years learning the local vernacular. In actual fact, the investigator may find that the potential subjects are only partially bilingual or that only some of the subjects are bilingual. In the first case, the hometown test serves as a check on the bilingual abilities of a subject. If he can perform to satisfaction on the hometown test, then his bilingual proficiency is not at issue. In the second case of only a portion of the potential subjects being bilingual, the investigator must decide if the bilinguals offer a good sample of the population or if they do not. For instance, Gillian Sankoff's study of multilingualism among the Buang of Papua New Guinea shows that the men understand New Guinea Pidgin significantly better than the women, but that on tests for other dialects of Buang men and women do not differ significantly (1968:169). This is evidence that among Buang dialects a common language approach using Pidgin which tested only men would not bias the results by leaving out women.

When the bilingual subjects are illiterate, then a common language oral approach is optimal. When the subjects are literate, then a common language written approach is optimal. The written approach is optimal in this case since the test booklets need be prepared only once. In the tests it is then possible to test a whole group of subjects individually in the same amount of time that one subject or one group collectively can be tested by an oral approach.

The choice between testing subjects as groups or as individuals may be influenced by who the subjects are, in particular by what their culture is like. In American culture, for instance, individualism is stressed and individuals do not know most of the people that are near them on any given day. In Melanesian cultures, however, the group is stressed and everyone in the village knows everyone else. My wife and I found that a method of group testing in Melanesia was more in tune with the culture. Whenever we entered a village a large group of people gathered around us. To isolate an individual subject with earphones while the remaining subjects waited their turn never seemed quite right. Sankoff reports the same kind of situation; however, she developed a strategy by which individual testing became appropriate (1968:177-8). After arriving in a village she chatted with the welcoming group for a while but then explained that she was tired from the walk to the village and asked to be excused so that she could rest. Upon arising most people were out of the village at work in their gardens. Thus it was possible for her to walk through the village and find some people to interview and test in relative privacy.

2.3.4 Summary of optimal methods

The discussion in the last section concluded that for monolingual subjects, the optimal test method was a vernacular language oral method. Of the methods listed in

Figure 2.1, Casad's and Sankoff's methods are the only ones which are appropriate. For bilingual subjects that are illiterate, the discussion indicated that common language oral approaches are optimal. The early methods (Section 2.2.1) and the method suggested in Section 2.1 fit this designation. Another alternative would be to modify Casad's or Sankoff's method to use common language questions (L. Simons 1977:241). For bilingual subjects that are also literate, the discussion indicated that a common language written approach is optimal. The methods of Yamagiwa, Bender and Cooper, and Ladefoged (Section 2.2.4) are appropriate here.

One aspect of the definition of optimality was time and effort. A method which yields the greatest amount of information with the least expenditure of time and effort is optimal. A deterrent to conducting an intelligibility survey or to completing one that has been started (of the 20 surveys Casad lists as having been conducted in Mexico, only 5 are listed as having been 100% completed, 1974:162) is the time required to conduct the sruvey. For the common language methods, the time required has pretty well been brought down to a minimum. In Section 2.1.5, I showed that the method I used in the Solomon Islands required one hour with an informant to collect a text and another hour alone to prepare the test tape. The early methods described in Section 2.2.1 might have taken a little longer since they made exact transcriptions and translations. The written methods with test booklets (Section 2.2.4) would require a little longer to prepare the booklets. For testing, one hour was required to test a group on a battery of test tapes. For the early methods in which the subject's translation was recorded, it would take no longer to administer the tests, though it would require additional time at a later date to listen to the recorded responses and score them. For the written methods, a whole classroom of school children were tested individually in the time I could administer the tests to one group collectively. For these methods, the time needed to prepare the test tape for a dialect is about two hours. Testing in one dialect could take as little as one hour when group testing is used.

In contrast to these, Casad's method for testing monolingual subjects requires two days per dialect by a two man team (Section 2.2.3). The preparation of the test tapes takes the first day, testing ten subjects takes up the second day. The method of preparing a test tape begins in the same manner as the other methods by eliciting a text and transcribing it. What takes so much longer is translating the set of questions that go with the test into each of the local dialects in which the test will be administered. After questions are translated they must be checked for accuracy with a pre-test panel and then dubbed into the test tapes. In addition an introductory tape is translated into the local dialect. In other words, a test tape must be redone for every dialect in which it will be tested. In the common language approaches a single test tape is made once and for all. This is the essential difference which makes test preparation require only two hours by a one person team as against one day by a two person team.

Sankoff's method of testing orally in the vernacular would require only a few hours by a one person team to prepare a test tape. This is because she did all the questioning personally rather than recording the questions on the test tapes. Thus she made test tapes once and for all rather than remaking them for each dialect. Although this method is optimal in terms of test preparation time, it has a major drawback in another sense. It requires several months, or longer, of preparation time spent in language learning for the investigator to achieve sufficient facility in the various local dialects to do all questioning personally.

I am aware of two methods for testing monolinguals in the vernacular which may help here. They have not received widespread attention in the literature, but they may reduce the time scale for vernacular intelligibility tests to a level comparable to that for the common language tests. They will do so in two ways: only a few hours would be required by one person to prepare test tapes, and learning of the vernacular would not be necessary. One could argue that these methods would yield results that were not as precise. However, I have already argued that the results of methods which yield percentage scores are already too precise for the level of statistical significance that can be attached to the results.

The first method is a sentence repeat method which was tried by Crawford in a pilot intelligibility survey in Mexico. It was abandoned in favor of a content repeat test which was subsequently developed into Casad's method for testing intelligibility. The sentence repeat test was as follows (Casad 1974:60). A free text was elicited for the basis of the test. Every third sentence out of a portion of this text was extracted and played back to a subject one at a time. The subject was asked to repeat the sentence. Crawford evaluated the responses on a five-point scale. He observed that for highly intelligible dialects the sentence repeat became so easy that a subject's response seemed more like mimicry than a test of intelligibility (Casad 1974:61). However, this need not be viewed as a liability. It simply indicates that the test is not sensitive enough to distinguish between different degrees of high intelligibility. In most cases we do not need to do that anyway.

Crawford observed that the results of the sentence repeat test showed little correlation to the results of the content repeat test. It was therefore dropped in subsequent studies. Casad, however, has suggested that it might be reinstated (1974:88). He credits Gudschinsky as saying that recent research in psycholinguistics has demonstrated that ability to mimic sentences of a different dialect is dependent on one's knowledge of both the grammatical structure and the phonological structure of that dialect. From my perspective, a great advantage of this kind of test is that the test tape can be constructed very easily and the one tape will then serve for all tests on that dialect.

A sentence repeat method could probably also be used by a survey technician who is a good phonetician but not a speaker of any local vernacular. He could rely on bilinguals in the village or on a bilingual traveling companion to explain how the testing would work. In scoring responses he would use the clues of immediacy of response, speed and timing of response, pitch contour of response, and phonetic similarity to the segmental phonemes of the utterance. A phonetic transcription of the utterances could serve as a standard against which to score.

A second simple method for testing intelligibility among monolinguals has been used by Robert Conrad in the Sepik region of Papua New Guinea (L. Simons 1977:250):

> This test consists of a number of simple questions such as, "Where is your father?", "Who is your brother?", and "How far away is your garden?" To construct a test tape a series of such questions is translated into the dialect of the reference point and recorded on tape. A test is administered by playing the questions one at a time to an informant at the test point. The subject is permitted to respond in whatever way seems most natural to him. If the subject answers the question, an appropriate response is taken to indicate understanding of the question. If, on the other hand, the subject prefers to translate the question, his translation is scored as correct or incorrect. The percentage of questions to which the subject gives an appropriate response is the measure of intelligibility.

Again, this method of testing requires that the test tape be created just once for all dialects. This method gives an interesting twist to the question approach. In the other question approaches, the subject hears a portion of text in the test dialect and is then asked a question about it in a language which he is sure to understand. In this approach, the text and the question are one and the same. Administering tests with this method would be bound to take half as long as in a text and question approach. With only the questions in the test, there is only half as much to play back. This method and the sentence repeat method deserve serious consideration as alternative methods to testing intelligibility among monolinguals.

CHAPTER 3

FINDING CENTERS OF COMMUNICATION

The previous chapter presented a number of ways in which intelligibility between dialects can be measured. This chapter tells what to do next: examine the patterns of communication to find groupings of dialects which can be served by common vernacular language programs. This chapter offers practical suggestions on how to apply the results of intelligibility testing to language planning in an area. The analysis techniques presented offer answers to the questions of how many vernacular language programs are needed in an area, and where those programs should be centered.

A vernacular language program is defined as any program which seeks to disseminate information by means of the vernacular language of a specific region. The medium of communication can be broadcasting, tape recordings, word of mouth, or literature. Literature programs are probably the most common. Materials produced in a vernacular literature program might include curricula and text books for primary and secondary education, translations of the Bible and liturgical materials by the church, or general and cultural reading materials for adult education. Such projects can be costly in terms of both money and effort. The strategy of the methods presented in this chapter is to find solutions which involve the least possible cost.

Basically, the problem is one of grouping together dialects which can be served by the same vernacular language program. Section 3.1 discusses the main criteria for making such groupings: adequacy and least cost. Section 3.2 presents a simple inspection method which can be used to find groupings of dialects which fit the adequacy and least cost criteria. Section 3.3 gives a step by step description of the grouping algorithm which could be translated into a computer program. Finally, in Section 3.4, a similar method developed by Joseph Grimes is reviewed.

3.1 The criteria of adequacy and least cost

Many of the developing nations of the world face the difficult challenge of trying to communicate with a multilingual population, a population which may include well over a hundred dialects or languages. Even among nations where a national language is firmly established, gross dialect variations of the national language and pockets of minority languages still exist. It may not be thought feasible for a country to initiate vernacular language programs in every one of its languages and dialects. On the other hand, if that country wishes to reach all of its citizens, it must carry out its programs in languages that are both understood and accepted by all groups concerned. Fortunately, communicating with every citizen does not usually require a language program in every dialect. Intelligibility tests, such as those described in Chapter 2, show where communication can take place across dialect boundaries. The need then is for criteria by which we can join dialects into larger groupings that can be served by a single vernacular language program. The two criteria suggested are adequacy and least cost (Stolzfus 1974:58-60).

The task is to define groups of dialects such that all dialects within a group can be served by a single language program, centered in one of these dialects. Since communication is the purpose behind the language program, a possible criterion for grouping is that all the dialects in the group understand the central dialect. Intelligibility itself is not a strong enough criterion, however, since it can span such a wide range of degrees. The _criterion of adequacy_ states that a dialect can be grouped with a central dialect if and only if its speakers understand the central dialect at a level which is deemed adequate for the intended purpose. Note that the level of adequacy is not fixed; it depends on the nature of the information to be communicated. For instance, if the purpose of the program were to broadcast news of current events, then the hearers would not be required to have as deep a command of the central dialect as if the purpose were to communicate about emotions, morality, or eternal values in a program of religious instruction. Note also that the criterion of adequacy says nothing about mutual understanding, but only about one-way understanding. That is, for a dialect to be grouped with a central dialect, it matters only that the former dialect adequately understand the central dialect. The degree to which the central dialect understands the other one is not relevant.

The adequacy criterion, when used to interpret intelligibility relations in a region, will designate a number of possible central dialects and a number of possible groupings around them. By itself it is not strong enough to suggest the best grouping among the possible solutions. To do this, the second criterion is added. The _criterion of least cost_ states that the best grouping is one which minimizes the number of central dialects. A major deterrent to vernacular language programs is the cost involved in studying the dialect, writing or translating the materials to be communicated, and then printing, recording, or broadcasting them. The total cost of vernacular language programs in an area is proportional to the number of central dialects in which specific programs are carried out. Thus if the grouping of dialects which requires setting up the least possible number of language programs is found, the least costly solution is normally also found. If the two criteria of adequacy and least cost are applied together, then such groupings will be found. The remainder of this chapter tells how this can be done.

3.2 An inspection method for analyzing patterns of communication

Patterns of communication can be diagrammed by drawing arrows onto a map of the test area. Simply by inspecting the pattern of arrows on the map, it is often possible to see a least cost solution which fits the given pattern. The method is basically this: (1) draw the patterns of communication on a map by representing each relationship of adequate understanding as an arrow from hearer to speaker, (2) find the dialect which is understood by the greatest number of other dialects and designate it as a center, (3) draw a loop which encloses all dialects that are reached by (that is, point to) that central dialect, (4) for all dialects remaining outside the loop, repeat the process beginning at step 2 and continue until all dialects are accounted for.

The procedure is now illustrated with sample data from Santa Isabel in the Solomon Islands. Seven dialects are spoken on Santa Isabel (Whiteman and Simons 1978; the data are adapted from Table 4). These dialects are, from northwest to southeast: Zabana, Kokota, Zazao, Blablanga, Maringe, Gao, and Bugotu. The communication patterns are set out in Figure 3.1. The dialects listed along the left hand side of the table are those of the hearers while the dialects listed along the top of the table are those of the speakers. A "yes" in the body of the table indicates that the given group of hearers understands the dialect of the speakers.

In fact, in this case they claim to have a command of the dialect which allows them to speak as well as hear it when communicating with someone from that region. This is defined as the level of adequacy for this analysis. A "no" indicates that the hearers do not understand the speakers at that level of adequacy.

Figure 3.1 Intelligibility on Santa Isabel Island

		Dialect of speaker:						
		ZAB	KOK	ZAZ	BLA	MAR	GAO	BUG
Dialect of hearer:	Zabana	yes	no	no	no	no	no	no
	Kokota	yes	yes	yes	yes	yes	no	no
	Zazao	yes	no	yes	no	yes	no	no
	Blablanga	no	no	no	yes	yes	no	no
	Maringe	no	no	no	no	yes	no	no
	Gao	no	no	no	no	yes	yes	yes
	Bugotu	no	no	no	no	no	no	yes

In the first step of the process, the patterns of communication are drawn onto a map. In this map all the instances of "yes" in Figure 3.1 are represented by an arrow pointing from the hearers to the speakers. This map is shown in Figure 3.2.

The second step in the process is to find the dialect which is most widely understood. This is found by locating the dialect which has the most arrows pointing to it. In Santa Isabel this is Maringe (MAR).

The third step is to map the extendability of the dialect just selected as a center. First the central dialect is underlined to indicate that it is a center. Then a loop is drawn which encloses all dialects that can understand the central dialect, but excludes all that cannot. Figure 3.3 shows the state of the analysis thus far.

Finally the second and third steps are repeated for all of the dialects which remain ungrouped. In the Santa Isabel example, only two dialects remain, Zabana and Bugotu. No arrows lead away from either of these dialects. Therefore, the only way in which they can be reached by a vernacular language program is if these two dialects themselves are centers for such programs. Thus we conclude that two additional centers are required, one at Zabana and another at Bugotu. These two dialects are underlined in the map and loops drawn around them to show the extendability of their language programs.

For the final map of the least cost solution, all extraneous arrows can be omitted, that is, omit all arrows which do not point to a central dialect. Figure 3.4 gives the final least cost analysis for Santa Isabel. Note that the inclusion of three dialects is ambiguous. Kokota and Zazao could be part of either the Zabana or the Maringe program and Gao could be part of either the Maringe program or the Bugotu program.

Figure 3.2 Patterns of communication on Santa Isabel

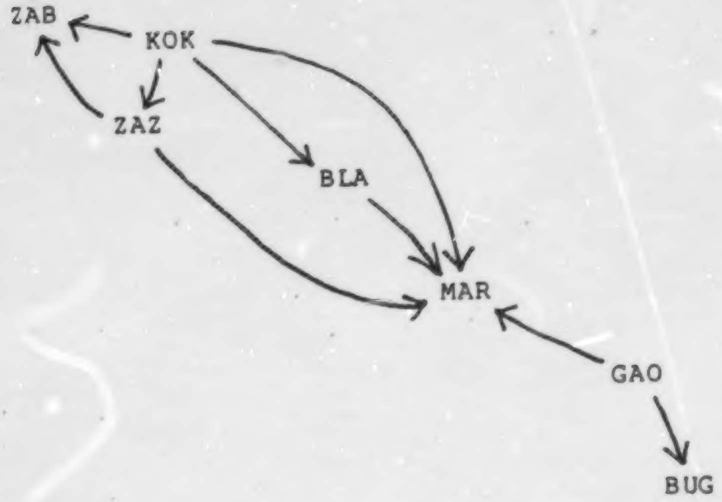

ZAB
KOK
ZAZ
BLA
MAR
GAO
BUG

Figure 3.3 The Maringe language program

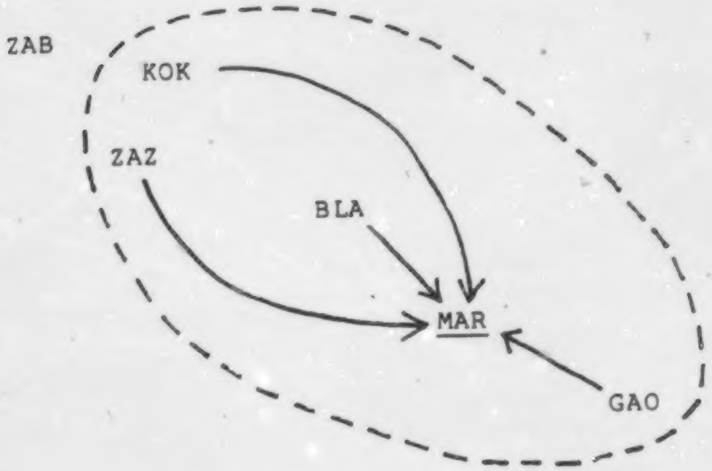
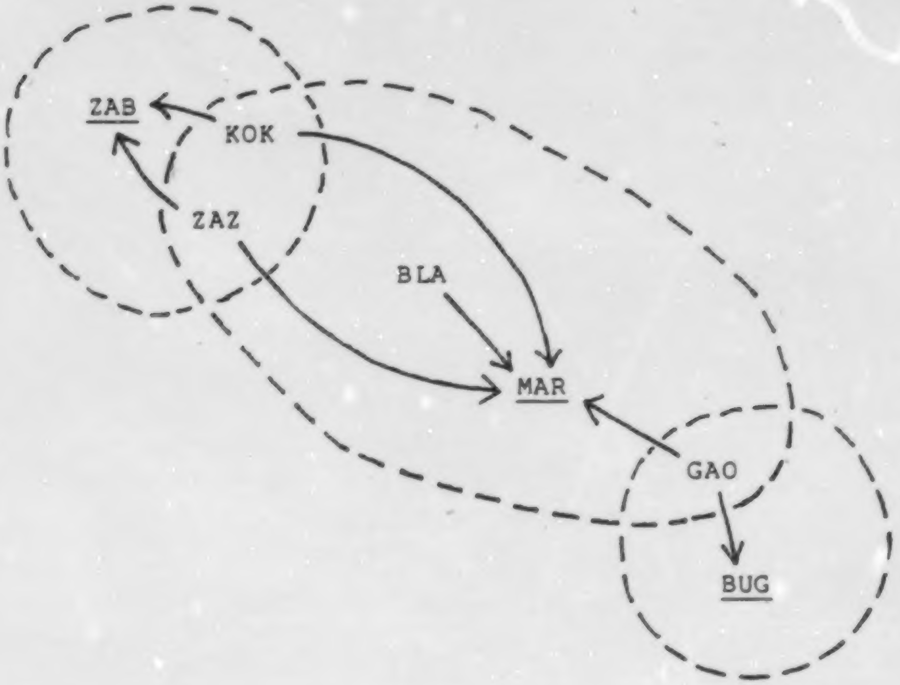

ZAB
KOK
ZAZ
BLA
MAR
GAO
BUG

Figure 3.4 Three vernacular language programs for Santa Isabel



Some data from the Northern Mixteco of Mexico are now analyzed to demonstrate an extension of the method. This is the analysis of data over successively lower levels of adequacy to develop a contour-like map of possible dialect groupings. The data are set out in Figure 3.5. The values in the table are percentages of intelligibility. The periods represent relations that were not measured. The table is taken from Grimes (1974:264) with three adaptations: the values in the table are percentages of intelligibility rather than intelligibility loss, the matrix is transposed, and three dialects (CC, CO, and AP) are omitted since they were not tested and have no effect on the grouping.

The patterns of communication are analyzed at successive levels of adequacy. First we might try 90% intelligibility as the level of adequacy. For the Northern Mixteco data, most of the hometown scores are not even 90%. There are no groupings at this level, so nine centers are required. Next, 80% intelligibility is taken as the level of adequacy. Any relations with 80% or more intelligibility are considered adequate, and any with less are not. Besides the hometown scores, only two relations are adequate at the 80% level, JE's understanding of CH and CS's understanding of CH. Thus at the 80% level, seven vernacular language programs would be required, one in CH to serve JE amd CS as well, and then one in each of the

Figure 3.5 Intelligibility in Northern Mixteco

		Dialect of speaker:								
		CZ	JE	CH	CS	CG	XB	CU	ZP	PT
	CZ	77	.	4	.	.	.	.	.	.
	JE	.	81	83	.	32	74	.	.	.
Dialect	CH	15	.	89	.	20	56	.	.	.
	CS	.	.	81	91	30	75	.	.	.
of	CG	.	.	21	.	81	19	.	.	.
	XB	.	.	78	.	17	84	.	.	.
hearers:	CU	66	.	23	.	.	.	86	.	.
	ZP	.	.	73	.	16	37	.	75	.
	PT	.	.	76	.	21	61	.	.	84

CZ = Cuyamecalco Zaragoza
CH = Santiago Chazumba
CG = Sta. Maria Chigmecatitlan
CU = Sta. Ana Cuauhtemoc
PT = Petlalcingo

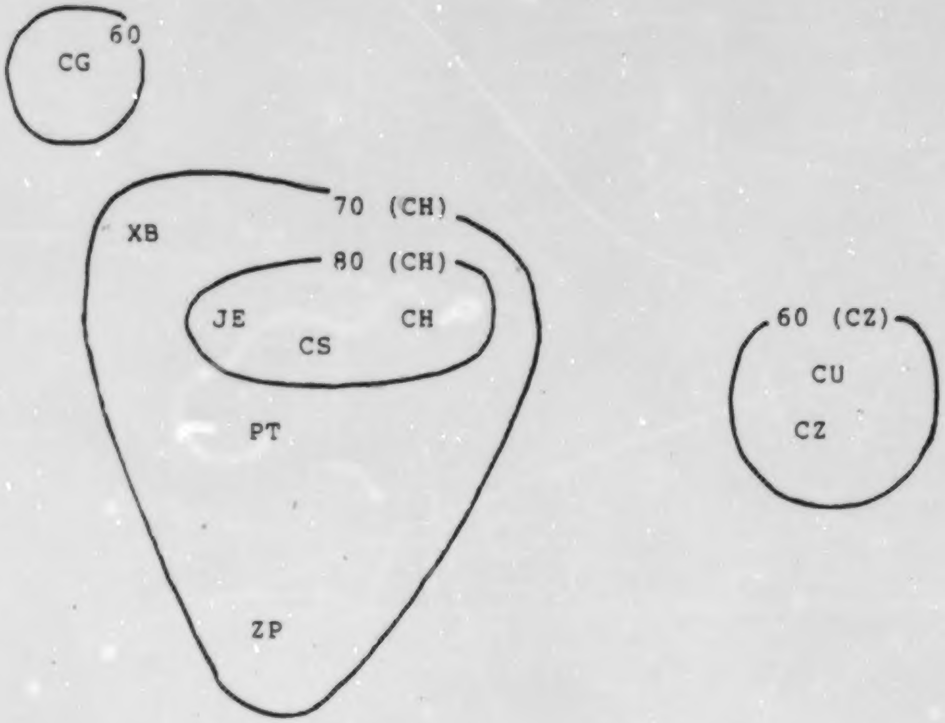
JE = San Jeronimo
CS = Cosoltepec
XB = Xayacatlan Bravo
ZP = Zapotitlan Palmas

other six dialects. When the level of adequacy is lowered to 70% intelligibility, the program at CH will extend to three more dialects, XB, PT, and ZP. The remaining three dialects still require their own programs. At the 60% level, a new group is possible: CU understands CZ at 66% intelligibility. Thus far, CG remains isolated. Only if we lowered the level of adequacy to 21% intelligibility would CG join in with the CH group. However, such a level of intelligibility is too low to conceive of as being very adequate, so the groupings are taken down only to the 60% level for the final presentation of results.

The results of the Northern Mixteco grouping are shown as a map in Figure 3.6. The loops showing the extendability of the dialect groups are shown as before. The only difference is that loops for different levels of adequacy are superimposed on the same map; the result is like a contour map. The loops are labeled with two items of information: the minimum percentage of intelligibility which is the level of adequacy for the enclosed group, and the name of the dialect which is the center. The labeling of loops by the dialect which is the center is an alternative to indicating centers by drawing arrows as was done in Figure 3.4. When loops are drawn at successive levels of adequacy, then arrows will cross loop lines and more than one arrow from a dialect may be required since a dialect can shift to a new center at lower levels of adequacy. When relations become complex, labeling loops is a clearer way to indicate centers than drawing arrows.

I have chosen to simplify the map by drawing only the loops which establish more inclusive groupings. A complete contour display would draw a loop around each dialect or group for each level. For instance, the CG dialect would be surrounded by four concentric circles, one for each of the intelligibility levels of 90%, 80%, 70%, and 60%. The large CH group would have two loops around it for 70% and 60%.

Figure 3.6 Dialect groupings in Northern Mixteco



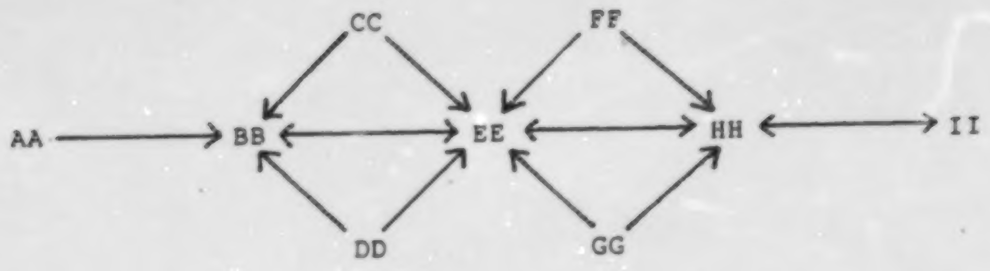
When all of the contour lines are drawn in, the relative distance between dialects can be found by counting the number of contour lines (divided by two) which separate them.

A hypothetical set of data typical of a dialect chain is now presented as a warning that the simple procedure described in the first paragraph of this section will not always yield a least cost solution. (In Section 3.3 a more complex procedure which always does is presented.) In Figure 3.7a the patterns of communication for the hypothetical data are shown as arrows. Figure 3.7b shows the first solution one is likely to arrive at by following the simple procedure: dialect EE in Figure 3.7a has the most arrows pointing to it so we designate it as a center and draw a loop. Only two dialects remain outside the loop, AA and II, and they do not understand each other so we set each of them up as the centers for separate language programs. This result with three language programs is shown in Figure 3.7b.

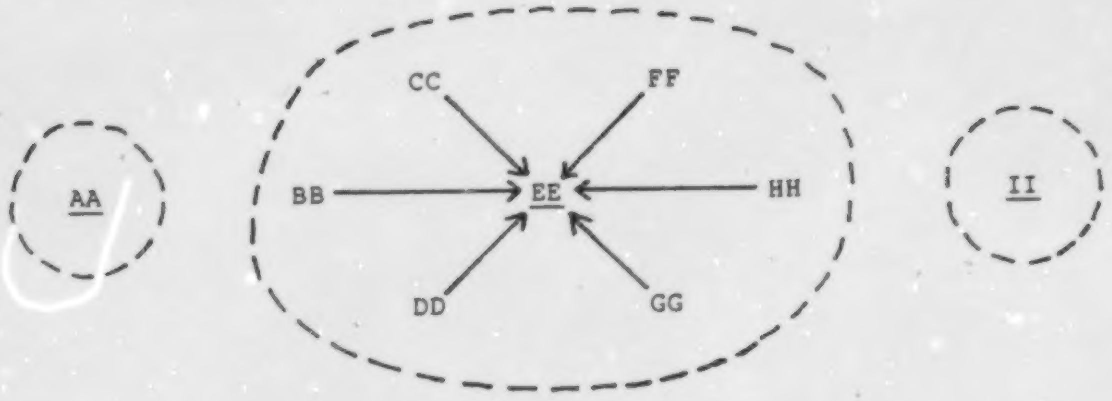
This result is not the least cost solution, however. Figure 3.7c shows that if dialects BB and HH are made centers, then all the dialects are reached with only two centers. In Figure 3.7b we went wrong by assuming that the dialect with the greatest number of arrows pointing to it had to be a center. Thus step 2 in the procedure, "find the dialect which is understood by the greatest number of dialects

Figure 3.7 Hypothetical data

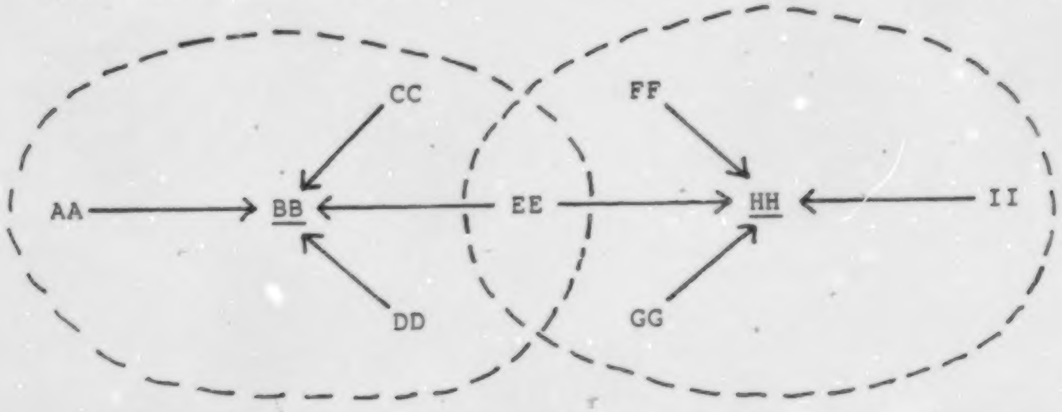
(a) Patterns of communication


CC
FF
AA
BB
EE
HH
II
DD
GG

(b) First solution


CC
FF
AA
BB
EE
HH
II
DD
GG

(c) Least cost solution


CC
FF
AA
BB
EE
HH
II
DD
GG

and designate it as a center," is not foolproof. However, it does turn out to be a handy rule of thumb which usually works. The least cost solution can be found by tracing the arrows which lead from the possible centers that were posited in the second pass over the data. In the first pass, EE is posited as a center. AA and II remain. Rather than just accepting AA and II as centers, we must see what points could serve them as centers as well. This line of inquiry points straight to dialects BB and HH and from Figure 3.7a the investigator can see that all dialects understand one of those two dialects.

3.3 An algorithm for finding all possible least cost dialect groupings

The procedure presented in Section 3.2 is simple and works well when there are not many dialects involved. However, if the investigator is not careful to trace out all the alternatives it may not yield the least cost solution, as was shown with the hypothetical data in Figure 3.7. As the number of dialects increases and the complexity of the pattern of arrows increases, this possibility becomes more likely. In this section, an algorithm for finding all possible least cost dialect groupings is presented. The algorithm is written in a prose format. However, it could be translated directly into a computer program which would determine least cost groupings automatically.

3.3.1 The least cost grouping algorithm

The algorithm is listed in Figure 3.8. It is to be read as a series of ordered steps. After each step is completed, the next step in the sequence should be performed unless there is a specific instruction to go to another step. Each of the steps is now discussed in turn.

(1) The only input data to the algorithm is the matrix of intelligibility relations as measured by testing methods described in Chapter 2. The algorithm is repeated for different levels of adequacy. First a level of adequacy must be selected. Then the matrix of intelligibility relations is transformed into an adequacy matrix: all intelligibility relations which are of an adequate level become 1's in the adequacy matrix and those relations which are inadequate become 0's. If there are values of intelligibility which were not measured and which have not been estimated by means of a predicting model (Section 6.2), then these also must be recorded as 0's in the adequacy matrix. The matrix of intelligibility relations from Santa Isabel in Figure 3.1 is already an adequacy matrix: "yes" is equivalent to 1 and "no" to 0. The intelligibility matrix for Northerm Mixteco (Figure 3.5) requires transformation. Groupings were computed at four different levels of adequacy: 60%, 70%, 80%, and 90%. For each adequacy level, a separate adequacy matrix must be computed. For instance, if the 70% level were being computed, all values of intelligibility greater than or equal to 70% would become 1's in the adequacy matrix, and all values less than 70% or missing would become 0's.

(2) Two variables are maintained during the algorithm. The first, n, is set equal to the number of dialects which are speakers in the intelligibility and adequacy matrices. The second, c, represents the number of centers in the solutions which are currently being tried. Initially this is set to one. The strategy is this: first all possible solutions with one center are tried. If not all the dialects can understand any one of the dialects adequately, then all the solutions with two centers are tried. If no two dialects can adequately reach all the

Figure 3.8 Least cost grouping algorithm

(1) Select a level of adequacy, then transform the intelligibility matrix into an adequacy matrix as follows:

Set all adequate values to one.
Set all inadequate and missing values to zero.

(2) Set n = the number of dialects. Set c = 1 (the lowest possible number of centers).

(3) Try all possible solutions with c centers. These possible solutions are all of the possible combinations of the n dialects taken c at a time. The number of such combinations equals $n!/(c!(n-c)!)$.

(3a) Test each possible solution by taking the logical *or* of the adequacy matrix vectors for the c centers as speakers.

(3b) If the logical *or* contains no zeros, then all dialects understand at least one of the c centers at an adequate level. Write down this solution (but keep looking as there may be more).

(3c) Return to step 3a and test another possibility until all possible combinations of c centers are exhausted.

(4) If any solutions were found, go to step 6. Otherwise, add 1 to c.

(5) If $c = n$, then go to step 6; the solution is that each dialect must have its own program. Otherwise, go to step 3.

(6) The least cost solution (or solutions) for the given level of adequacy has been found. If all desired levels of adequacy have been analyzed, then quit. Otherwise, go to step 1.

dialects, then all possible solutions with three centers are tried. This continues until a solution is found. Since the search begins with the least possible number of centers, one, and works up, the solutions reached are guaranteed to be the ones involving the least possible number of centers.

(3) The third step is a complex step made up of three substeps. This step is the heart of the algorithm; in this step the possible solutions are tested to find the least cost solutions. A possible solution with *c* centers is any combination of *c* dialects. The total number of such possible solutions is the number of possible combinations of the *n* dialects taken *c* at a time. This number is defined by the quantity $n!/(c!(n-c)!)$, where *n*! is read *n*-factorial and equals the product $(n)(n-1)\ldots(2)(1)$. For instance, the total number of combinations of 7 dialects taken 3 at a time is:

$$7!/(3!(7-3)!) = \frac{(7)(6)(5)(4)(3)(2)(1)}{(3)(2)(1)(4)(3)(2)(1)}$$

$$= \frac{(7)(6)(5)}{(3)(2)(1)} = 35$$

Thus for 7 dialects there are 35 possible combinations of 3 dialects that could serve as centers.

For an example, all the possible solutions when there are four dialects, called A, B, C, and D, can easily be enumerated. In the list which follows, the braces enclose sets of dialects which serve as centers. Each set is a possible solution. Note that the ordering of the dialects within the sets is immaterial:

Solutions with one center = 4 possibilities:
  {A}; {B}; {C}; or {D}
Solutions with two centers = 6 possibilities:
  {A,B}; {A,C}; {A,D}; {B,C}; {B,D}; or {C,D}
Solutions with three centers = 4 possibilities:
  {A,B,C}; {A,B,D}; {A,C,D}; or {B,C,D}
Solutions with four centers = 1 possibility:
  {A,B,C,D}

(3a) Each possible solution is tested by taking the logical *or* of the adequacy matrix vectors for the *c* centers as speakers. In the matrix for Santa Isabel, Figure 3.1, the vectors for the centers as speakers are the columns; the column for a dialect tells which other dialects understand that dialect adequately. An acceptable solution is one in which each dialect understands at least one of the centers adequately. An easy way to determine if all of the dialects understand at least one center is to compute the logical *or* of the speaker vectors. The operation of logical *or* yields zero if all its operands are zero; it yields one if at least one of the operands is one. The logical *or* of the speaker vectors for the three dialects which comprise the least cost solution for Santa Isabel is as follows:

ZAB	MAR	BUG	Logical Or
1	0	0	1
1	1	0	1
1	1	0	1
0	1	0	1
0	1	0	1
0	1	1	1
0	0	1	1

The fact that all dialects understand at least one of the centers is indicated by the fact that all elements in the result vector are ones.

Any other combination of three dialects on Santa Isabel yields an unacceptable solution. For instance, the combination of ZAZ, BLA, and BUG leaves out two dialects:

ZAZ	BLA	BUG	Logical Or
0	0	0	0
1	1	0	1
1	0	0	1
0	1	0	1
0	0	0	0
0	0	1	1
0	0	1	1

(3b) If the logical *or* vector contains no zeros, then all dialects understand at least one of the centers at an adequate level. That set of *c* centers is therefore an acceptable solution to be written down. One should not stop here, however. All of the remaining possibilities with *c* centers must be checked to see if there are other solutions. If the logical *or* vector does contain a zero, however, then proceed without recording anything. Note that as long as the acceptability criterion is stated as a result vector that contains no zeros (rather than one that contains all ones), the operation of addition can be used as readily as the logical *or*. In that case, the result vector would tell how many of the centers each dialect understands adequately.

(3c) Return to step 3a and test another possible solution until all possible combinations of *c* centers are exhausted.

(4) If any acceptable solutions were recorded in step 3b, then all of the least cost solutions have been found for that level of adequacy. Jump to step 6. Otherwise, add 1 to *c* in order to search for possible solutions containing one additional center.

(5) If *c* equals *n*, the number of dialects, then the only solution is that each dialect must have its own language program for that level of adequacy; proceed to step 6. Otherwise, go back to step 3 and test all possible solutions with the increased number of centers.

(6) The least cost solution (or solutions) for the given level of adequacy has been found. If all desired levels of adequacy have been analyzed, then quit the

procedure. Otherwise, go to step 1, select a new level of adequacy, and repeat the procedure.

3.3.2 Deciding among multiple least cost solutions

When more than one solution with a minimal number of centers is found, there are at least four strategies that can be used to decide among possible solutions. Each of the strategies involves applying a principle of least cost in some other sense.

One strategy is to compare competing solutions for overall information loss. Information loss is defined as the complement of intelligibility. Thus if intelligibility is 85%, then the information loss is 15%. The total information loss for each solution is computed. The solution which results in the least overall information loss, costs the least in that respect. In computing total information loss, be sure that each dialect is counted only once; if a dialect understands more than one center adequately, then group it with the center which it understands best. This will minimize the information loss. The computation of information loss can be refined by computing the average information loss per individual in the region. In this way large dialects will carry more weight in the computations than small ones. To compute the average information loss per individual, sum the product of information loss times population for each dialect. Then divide by the total population of the region.

A second criterion is logistic cost. Establishing a center for a vernacular language program requires the transportation of personnel, equipment, and supplies from an administrative headquarters to the dialect area. A logistic cost could be assigned to each center. For instance, the center which was the most difficult and expensive to travel to would be the most costly. The possible solutions could then be compared for logistic cost by summing the costs of their individual centers.

A third dimension of least cost is sociocultural. Centers can be defined in terms other than intelligibility adequacy. In Section 6.1.4 the social side of centers in dialect systems is discussed. For instance, geography, population, economy, and politics can define centers. So can linguistic similarity. In Section 6.3.2 many of the different criteria which define the center of the Santa Cruz dialect system are listed. Ideally, the centers revealed by the analysis of communication patterns should coincide with the sociocultural centers in the region. If they do not, the materials which eminate from the language program may not be accepted by the people. Thus the solution which best fits the pattern of sociocultural centers may cost the least in terms of unacceptability.

A fourth aspect of least cost is stability of groupings. It is possible that groupings at different levels of adequacy will be used in the same language program. For instance, written materials for the beginning reader should be as similar as possible to his hometown dialect, whereas experienced readers can tolerate more variation and use literature with a wider extendability. If groupings are stable over many levels of adequacy and if the loops on a map are always concentric rather than crisscrossing, then the groupings lend themselves well to a strategy of hierarchical inclusions for materials at different levels. If centers shift and dialects regroup for different levels, then preparation of materials at different levels is more costly. When evaluating competing solutions for a given level of adequacy, the solutions for the level above and below should be consulted.

3.3.3 Computational refinements to the algorithm

The algorithm presented in Section 3.3.1, because it will eventually try all possible solutions if necessary, can be a time consuming one. This section describes some shortcuts that reduce the time required to analyze matrices, especially by a computer program.

In the best case, that in which one center is adequate, only *n* solutions are tested (where *n* is the number of dialects). In the worst case, that in which every dialect requires its own language program, the number of solutions which are tested before finding this out is 2 to the *n*th power minus 2. If there are 5 dialects, then this number is only 30. If ther are 10 dialects, this number is 1022. If there are 20 dialects, over one million possible solutions have to be tested -- 1,048,574 to be exact. Clearly something needs to be done to prevent testing impossible solutions (which most of them are).

Taking the logical *or* of a set of vectors is the easiest way to find out if a solution is acceptable; however, in most cases there is an easier way to find out if a solution is not acceptable. We could save many fruitless vector operations by first making a simple test to see if the current solution is even possible. Plausibility can be measured by noting the total number of 1's in the vectors being considered. If the number of 1's is less than *n*, then those dialects as centers could not possibly be an adequate solution. The advantage of using the number of 1's is that it need not be counted every time; rather, the number of 1's in a vector can be counted once and for all at the beginning and stored with the vector. To test the plausibility of a solution involving a set of *c* possible centers, the counts for those *c* vectors are summed. If the total is less than *n*, then the vectors are not *or*ed. It is much faster to sum *c* numbers, than to *or* *c* vectors of length *n*.

These refinements can be added to the algorithm in Figure 3.8 as follows: A new instruction is added to the end of step 1, "Count and store the number of 1's in each vector for the dialects as speakers." Step 3a becomes, "Test the plausibility of the solution by summing the counts for the *c* vectors. If less than *n*, then go to 3c. Otherwise, take the logical *or* of the vectors."

We can take this refinement even further. If in step 1 the matrix vectors are rearranged in the order of the counts for the vectors, then it is possible to know that when the current combination fails, certain of the remaining combinations will also fail. For instance, when testing for one center, if the first vector fails the plausibility test, then so will all remaining vectors since they have an equal or fewer number of 1's; it is possible to jump directly to testing the possible solutions with two centers. Likewise, if the sum of counts for the first two vectors fails the plausibility test, then processing can proceed straight to three-center possibilities.

The complexities in this refinement come in determining what to do after a plausibility test fails. For instance, if two-center solutions are being tested and the first combination to fail the plausibility test is the first with the fifth vector, then there is no use testing any more solutions with the first vector. However, it is necessary to begin testing solutions with the second vector. It is still possible that two with three, two with four, and three with four would be solutions, but never beyond the fourth. The method is to back up and advance the vector preceding the one which failed. When all the vectors are adjacent, and the plausibility test fails, then no more solutions for that many centers need be tested. An algorithm for just this aspect of the plausibility testing is beyond the scope of this discussion.

These refinements speed up the computation of least cost solutions, but it is not yet clear just how much. Earlier it was stated that the total number of possible solutions to test in the worst case of each dialect as its own center is 2 to the nth power minus 2. For large values of n this number takes on astronomical proportions. However, this holds only for the algorithm in Figure 3.8. When the algorithm is refined to order the vectors for the number of 1's they contain and to make a plausibility test, only n plausibility tests will be made in the worst case, because all vectors contain only one 1. The first plausibility test for every value of c less than n will fail. Thus in the refined algorithm, the best case of one center and the worst case of n centers, require a processing time on the order of n. For cases in between, more processing is required. At this point I do not know what the maximum and average processing times for the refined algorithm are.

3.4 Grimes's optimization method for grouping dialects

The ideas and methods presented in the first three sections of this chapter grow directly out of Joseph Grimes's work on "Dialects as Optimal Communication Networks" (1974). In this section I review his optimization method.

The optimization method is based on a principle of least cost. The method is widely used in the field of economics where the principle of least cost is well understood. A typical economic problem of this sort involves a manufacturer who needs to distribute his product to consumers in a wide geographical area. He would phrase the question of least cost something like this:

> In this geographical area, what is the most inexpensive way to supply every potential consumer with the product so as to assure greatest profits for the company?

For the manufacturer the most inexpensive approach could be one of several alternatives. It might be to have one central manufacturing plant and to distribute the products by truck. Or it might be less expensive to build small manufacturing plants in each of the cities where the product is to be distributed. The most inexpensive solution would probably involve a combination of assembly plants in primary centers with trucking to secondary centers. The configuration of the most economical solution is based on a compromise between the one-time cost of building factories, the cost of operating them, and the cost of trucking. The economist can assign a dollars and cents value to each possiblity in order to determine the solution which is the least expensive overall and will thus yield the greatest profits.

Grimes (1974) applied this principle of least cost to the analysis of patterns of communication. For the analysis of dialect groupings, he defined the question of least cost like this (1974:261):

> In a geographical or social area, what is the smallest set of speech communities such that adequate communication at a given threshold level can be established with every individual in the area by using the speech of at least one of the communities?

In Grimes's analogy to the economics problem, the cost of building and operating a factory corresponds to the cost of establishing a center for a vernacular language program; the cost of trucking corresponds to the cost of communicating with the other dialects. The cost of communicating is measured by the amount of information lost; the cost of establishing a center is controlled by a fixed cost value which is activated any time a dialect is a center. The fixed cost makes it uneconomical to use more than a minimal number of centers. The fixed cost is actually stepped through a series of values, called threshold levels, to yield a series of groupings at different levels of adequacy. When the threshold level (fixed cost) is low, then it is feasible to have many centers; when it is high, then only a few centers can be afforded.

In Figure 3.9, the algorithm for Grimes's optimization method is written out in a step by step prose format. Detailed instructions with examples on how to use the method are given in three sources: Casad 1974:36-45, Grimes 1974, and Arden Sanders 1977b. Therefore I will not repeat those detailed instructions and examples here. Rather, the listing of the algorithm serves as a point of reference for the evaluation of the method which now follows.

The optimization method has four hidden pitfalls which its user must be aware of: the interpretation of thresholds, the definition of least cost, the treatment of missing values, and degenerate solutions. The first two problems can be treated by reformulating the optimization method in the way I suggest in the following discussion. It should be noted that Grimes has accepted these suggestions and now uses the reformulated version of the optimization method. The least cost algorithm of Section 3.3 also avoids these problems. The third pitfall of missing values affects both the optimization method and my least cost method. The final problem of degenerate solutions is avoided by using the least cost algorithm.

(1) The interpretation of thresholds - The interpretation of the threshold levels has been incorrect. Grimes (1974:262) interpreted the thresholds as follows:

> For any communication effort [intelligibility loss] that is greater than the threshold level, the fixed-cost function renders it more economical to create another network than to add the test point concerned to an exisiting network. But for any communication effort that is not greater than the threshold level, the fixed-cost function renders it more economical to include the test point in an existing network than to create a new network with its own additional fixed cost.

Likewise, Casad (1974:46, 83ff.) speaks of an intelligibility threshold of 80% corresponding to a communication cost of 20. He suggests that 80% intelligibility is about the level of adequate intelligibility and thus that optimizations at the fixed cost level of 20 give groupings for the 80% level of adequacy. This is where the interpretation of thresholds goes astray -- there is not a one to one correspondence between fixed cost and adequacy or communication effort.

In the first place, the fixed cost, or threshold, value is sensitive to the differences between intelligibility measures, not to their absolute values. This is seen in the Northern Mixteco data (see Figure 3.5) which are optimized by Grimes (1974:265). At the threshold level of 10, the dialects JE, CH, CS, XB, ZP, and PT are assigned to the CH dialect as center. The communication efforts (or intelligibility loss) for these dialects with CH are 17, 11, 19, 22, 27, and 24, respectively. These correspond to intelligibility percentages of 83%, 89%, 81%, 78%, 73%, and 76%, none of which is greater than 90% as the interpretations of Grimes and Casad would suggest. In each case, the communication effort is greater

Figure 3.9 Grimes's optimization algorithm

(1) Transform the intelligibility matrix into a cost matrix by changing each intelligibility score to a measure of information loss. (85% intelligibility = 15% loss)

(2) Select the fixed cost (threshold) level.

(3) Initially assign each test point (dialect of hearer) to a center. The initial center is the reference point (dialect of speaker) which it best understands (lowest information loss). This is generally itself. Throughout the analysis, any reference point with at least one dialect assigned to it is a center; if no dialects are assigned to it, it is not a center.

(4) Step through the cost matrix comparing all possible pairs of reference point vectors. First compare the first with the second, the first with the third, and so on to the nth. Then compare the second with the third, with the fourth, and so on until all pairs are compared.

   (4a) Compute the cost for the two reference point vectors. If the first reference point is a center add in the fixed cost; if the second reference point is a center add in the fixed cost. For all the dialects assigned to either reference point, add in the information loss. If the cost is zero (neither dialect is a center) then repeat step 4a on the next pair of reference points. Otherwise, continue.

   (4b) Now try one of the following three things in an effort to minimize the cost for the two reference points: (1) take all dialects assigned to the second reference point and reassign them to the first one, (2) take all dialects assigned to the first reference point and reassign them to the second one, and (3) take all dialects assigned to either of the reference points and reassign them to the one which results in the lowest information loss. (When both reference points are centers, the first two options may reduce the cost by requiring one less center, while the third option may reduce it by minimizing information loss.)

   (4c) Recompute the cost for the two vectors for each of the three possibilities. If one of the three reassignments yields a lower cost than the original cost from step 4a, then shift the assignments to the least cost configuration. If there is a tie for the least cost, the first option has first priority, the second has next, and the third last.

   (4d) Go back to step 4a and process the next pair.

(5) During the whole pass through the matrix in step 4, if no assignments were shifted in step 4c, then go to step 6. Otherwise, return to step 4 and make another pass.

(6) The optimal (least cost) solution for the given threshold value has been found. If desired, go back to step 2 and optimize for another threshold. Otherwise, quit.

than the threshold level of 10, but the fixed-cost function finds it most economical to include the test points in an existing network. The error is in assuming that the threshold level is compared to the communication effort. It is not; it is compared to the difference in communication effort between two possible solutions.

The example given in Tables 2 and 3 of Grimes (1974:264-265) illustrate this point. The example is reproduced here in Figure 3.10. Before optimization, the XB dialect is assigned to itself as center (designated by the asterisk). The communication effort is 16. If XB were assigned to CH, the communication effort would increase to 22. 22 is greater than the threshold value of 10, so the interpretation in the above quotation would suggest that XB cannot be assigned to CH at this threshold level. However, increasing the communication effort from 16 to 22 is accompanied by a decrease of total fixed cost factors from 20 to 10, since one less center is required. The overall effect is a decrease in cost and thus the solution with one center is optimal for a threshold of 10, even though the intelligibility loss is 22. XB was joined to the existing network because the difference in communication costs was less than the threshold value.

Figure 3.10 Threshold corresponds to differences, not actual cost

a. Before regrouping

		Test points										
		CZ	JE	CH	CS	CG	XB	CU	ZP	PT	Fixed cost	Total cost
Reference point	CH	96	17*	11*	19	79	22	77	27	24	10	64
	XB	999	26	44	25	81	16*	999	63	39	10	

b. After regrouping

		Test points										
		CZ	JE	CH	CS	CG	XB	CU	ZP	PT	Fixed cost	Total cost
Reference points	CH	96	17*	11*	19	79	22*	77	27	24	10	60
	XB	999	26	44	25	81	16	999	63	39	0	

This quirk in the method does not appear in Casad's examples because in every case he uses matrices in which the raw scores are adjusted to raise hometown scores to 100% (a cost of 0). Therefore when a regrouping would shift a dialect from itself to another dialect as center, the difference in communication costs is the cost with the other dialect minus zero. In other words, in this special case, the threshold level does correspond directly to the communication cost. When raw intelligibility scores are optimized, and the threshold values are interpreted as corresponding directly to intelligibility levels then the raw scores are actually being subjected to an implicit constant adjustment for subject abilities (Section 5.2.4). That is, it is as though the difference between the hometown score for the subjects and 100% has been added to all intelligibility scores for that group of subjects.

Whether communication cost is based on raw or adjusted intelligibility scores, there will not be a correspondence between threshold level and intelligibility level when regroupings involve shifting more than one dialect at a time. In Figure 3.11 a hypothetical example is given. Such situations do arise in field data (Arden Sanders 1977:302 points out an example in the Mazatec data). However, the point is easier to see if a minimal example is constructed. The example shows two reference points (the speakers) and three test points (the hearers). The optimization is for the threshold level of 20. In Figure 3.11a, AA is the center for itself and BB is the center for BB and CC. Since two centers are involved and the communication cost of CC grouped with BB is 5, the total cost for this configuration is 45. Figure 3.11b shows the attempt to reduce the cost by using one center instead of two. To shift all the dialects to AA as the center looks plausible since the communication effort for BB with AA and for CC with AA is 15. This is less than the threshold level of 20. However, since two dialects are going to be regrouped this amounts to a total information loss of 30. The total cost including the fixed cost value is 50 and is higher than the solution using two centers. Therefore, all of the dialects would group together with one center at a level of 85% intelligibility, but not at a fixed cost threshold of 20.

The conclusion is that the threshold value does not correspond directly to the intelligibility level. It corresponds to the difference in summed communication cost for two possible solutions. Thus it is difficult to assign a meaning to threshold values which is both meaningful when applying results in the real world and is consistent.

(2) The definition of least cost - It is in the definition of least cost that Grimes's original analogy to the transport problem breaks down and leads to the misinterpretations just discussed. We saw this in the last example where the threshold level of 20 blocked the regrouping of two dialects with a communication cost of 15. The question we must ask is, "Are two fifteens worse than one twenty?" In economics, losing two fifteen dollar checks is certainly worse than losing one twenty dollar check. In the economic transport problem, the units by which the cost of building a factory and the cost of trucking goods are computed and compared are the same -- dollars and cents. This is what makes the optimization algorithm work. However, in the intelligibility analogy, the two kinds of cost are not comparable. Communication cost is measured in terms of information loss while establishing centers for vernacular language programs is measured in terms like personnel, transportation, equipment, and supplies. The analogy further breaks down when the meaning of information loss is examined. Is it worse for each of two people to lose 15% of the information in a message than it is for one person to lose 20% or 25%? I would think not.

The definitions of the criteria of adequacy and least cost I presented in Section 3.1 are the same as Grimes and Casad have in mind when they describe the optimization method. They define the problem as being one of finding the smallest possible set of centers (least cost criterion) capable of establishing communication at an adequate level (adequacy criterion) with the entire area (Grimes 1974:261, Casad 1974:37). In Figure 3.11 we saw that the optimization method does not

Figure 3.11 Threshold corresponds to sums, not individual costs

a. Before regrouping

		Test points			Fixed cost	Total cost
		AA	BB	BB		
Reference points	AA	0*	15	15	20	45
	BB	30	0*	5*	20	

b. After regrouping

		Test points			Fixed cost	Total cost
		AA	BB	CC		
Reference points	AA	0*	15*	15*	20	50
	BB	30	0	5	0	

actually do this, if we try to interpret the threshold levels in terms of levels of adequacy.

The optimization method can be reformulated as follows to find solutions which fit the two criteria of adequacy and least cost defined in Section 3.1. The reformulation is expressed as changes to the algorithm in Figure 3.9. In step 2, the threshold level becomes the level of adequacy. Fixed cost is given a different meaning in step 4. The adequacy level is used to determine if dialects can be shifted to a new center. If their understanding of the new center is adequate, they can; if it is not, they cannot. In step 4a, the total cost is defined in a different way. The fixed cost associated with establishing a center is an arbitrarily high constant at all levels of adequacy. It is so high that the sum of information loss for a center will never exceed it (_n_ times 100%, for instance). The total cost for two reference point vectors is then computed as before. In step 4b, for each of the three options, dialects which understand the potential new center at an adequate level (that is, information loss equal to or less than the threshold value) are shifted. Otherwise, dialects cannot be shifted. In step 4c, that of finding the least cost configuration for the two vectors, the arbitrarily high fixed cost ensures that a configuration with one center will always cost less than one with two centers. When comparing configurations with the same number of centers, the one with the least overall information loss costs less. The modifications then are these: the threshold equals level of adequacy, the fixed cost is an arbitrarily high constant, and dialects can shift to a new center only when the information loss is within the level of adequacy.

(3) Treatment of missing data values - Grimes lists one of the advantages of the optimization method of dialect grouping as being that it gives "useful results from matrices that can be filled in only partially" (1974:261). It is true that the method will give results from incomplete data, but using incomplete data can be hazardous to the unwary investigator. This is true not only of the optimization method but of the methods I presented in Sections 3.2 and 3.3 as well. It is important to understand the effects of missing data.

The grouping algorithms do not actually operate on matrices with holes in them. The investigator does actually fill in all the holes created by missing data. In the case of an adequacy matrix for the method of Section 3.3, missing values always transform to zeros. In the case of a cost matrix, missing values always transform to arbitrarily high amounts of information loss. The result is that when there is a missing value, it is never possible for the dialect of the hearers to group with the dialect of speakers for which it was not tested.

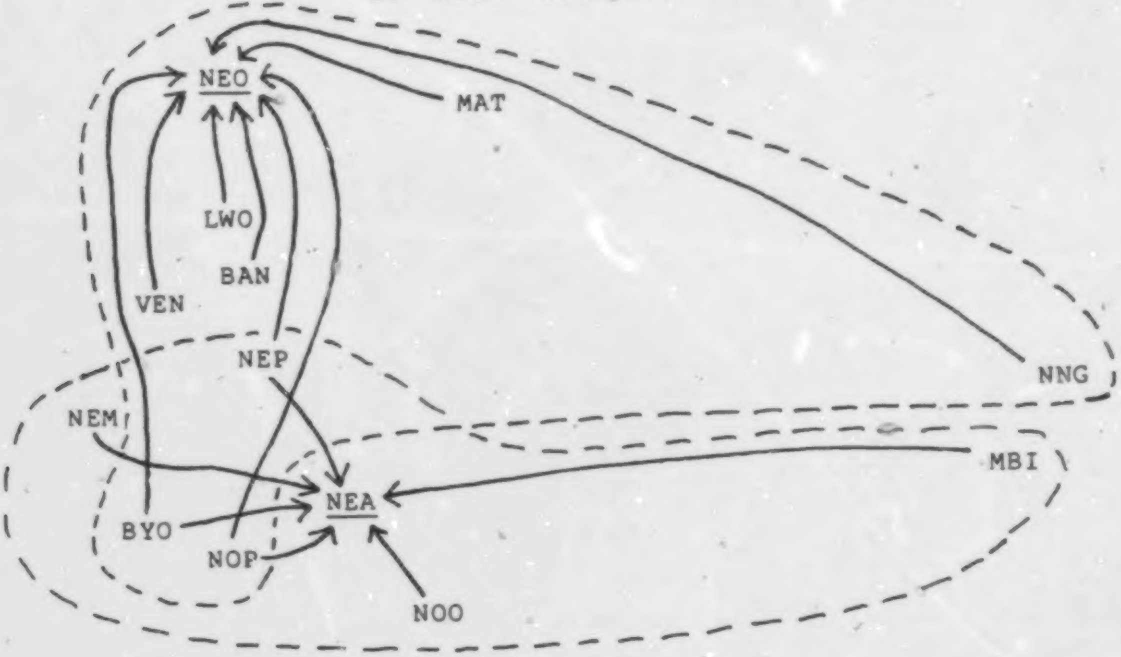
One effect of this is seen in matrices which are not square. Casad (1974:44-45) illustrates the optimization method on data from the Ocotlan Zapotec area of Mexico. In that intelligibility survey, 7 test tapes were used but they were tested in 10 dialects. The published matrix has 7 rows and 10 columns. Therefore, there are no hometown scores for three dialects. This means that those three dialects cannot even have themselves as a center; at least that is what a computer program which optimized the 7 by 10 matrix would assume. For one of these three dialects (San Andres, An, column 8) the lowest information loss in the matrix is 15% with Ayoquezco (row 5). This means that even at the zero threshold level, San Andres groups with Ayoquezco. Casad rectifies the situation in the dialect map (page 45) where San Andres remains isolated until it groups with Ayoquezco at the 15 threshold. However, he would not have gotten that result if he had strictly applied the optimization algorithm to the cost matrix on page 44.

The same example from Casad illustrates another effect of missing values. Since a dialect cannot group to a reference point for which it was not tested, there can be a grouping which includes all dialects if and only if there is a reference dialect on which all dialects were tested. In the Ocotlan Zapotec matrix there is no such reference dialect. Ocotlan is the main reference point with seven dialects tested on it. But the three dialects which were not tested on Ocotlan cannot group with Ocotlan. Actually if the matrix (page 44) were optimized up to the 100 threshold and beyond there would remain three disjoint dialect groups -- 2 and 6; 7, 9, and 10; and the other five dialects. In mapping the dialect network, Casad (page 45) estimated some missing values in order to allow the groupings to converge. For instance, the convergence of the DO-TI group with the IN-OC-MA group at the 26 threshold depends on the estimation that the missing value of intelligibility loss for DO on OC is equal to or less than 26.

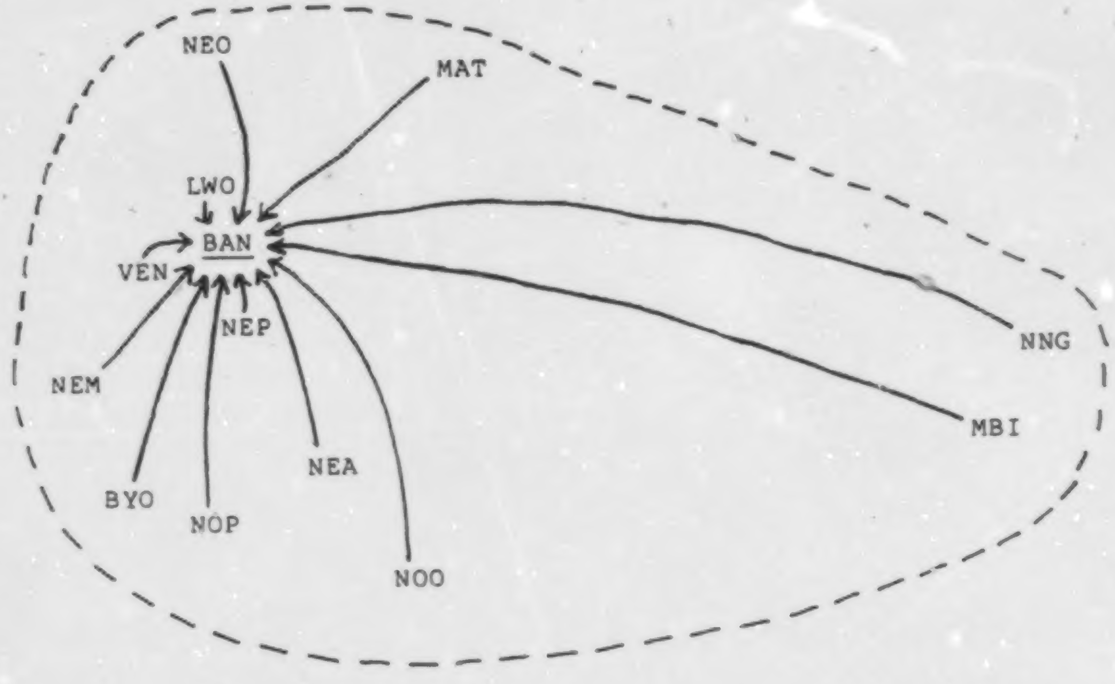
The hazards of unwittingly applying these grouping techniques on incomplete data are dramatized by an analysis of the results from the intelligibility survey of Santa Cruz Island. The intelligibility matrix is found in Table 2.2.6 of Appendix 2. In Figure 3.12 the least cost grouping technique of Section 3.3 is applied to these data. The top half of the figure contains a map showing the least cost solution on the incomplete data matrix. The level of adequacy is 3, or full intelligibility. The least cost solution calls for two centers, one at NEO and one at NEA. Even if the level of adequacy is lowered to 2, partial intelligibility, two centers are required. The intelligibility matrix is far from complete, however. When the survey was conducted, it was known that the dialects of LWO and BAN were the central ones on the island in terms of geography, population, and community

Figure 3.12 Dialect groupings on Santa Cruz Island

a. With incomplete data


NEO
MAT
LWO
BAN
VEN
NEP
NNG
NEM
MBI
NEA
BYO
NOP
NOO

b. With matrix completed by estimations


NEO
MAT
LWO
BAN
VEN
NEP
NNG
NEM
MBI
NEA
BYO
NOP
NOO

facilities (Sections 6.1.4 and 6.3.2). Informant opinions showed that everyone claimed to understand these central dialect (Appendix 2.1.8). There were thirteen dialects involved in the survey and it was not feasible to test all dialects against all others. (To be exact, 77 out of the 169 possibilities, or 46%, were tested.) The test tapes which were most frequently not used were those from the central dialects. When a group claimed full intelligibility with the central dialects and then scored full intelligibility on a dialect which was beyond the centers in the dialect chain (such as when southern dialects were tested on NEO) then intelligibility with the central dialects was a sure conclusion and not tested. Efforts were concentrated on results that could not be interpolated.

Thus it turns out that the central and most widely understood dialects were not actually tested for in the intelligibility tests. The result is that it is impossible to show the true least cost dialect grouping for the island from the measured intelligibility matrix. In Figure 3.12b a complete intelligibility matrix for Santa Cruz is analyzed by the same technique. The matrix was completed by estimating missing values (Appendix 2.1.11) with the predicting model developed in Chapter 6. The result is that one dialect, BAN, can serve the whole island as a center for a vernacular language program.

(4) Degenerate Solutions - One characteristic of the optimization method which has not yet been mentioned in the intelligibility literature is that it can lead to degenerate solutions. These are solutions which may not be unique. In step 4c of the algorithm (Figure 3.9) when two or more reconfigurations of two vectors lead to equal and optimal reductions in cost, the algorithm specifies that shifting all dialects to the first vector has top priority, shifting them to the second vector has next priority, and reshuffling them between the two vectors has lowest priority. This is where degeneracy can arise. The method always picks one of the optimal configurations and ignores the rest. It just may be that following the latter configuration would have led to another solution which was equally optimal.

The algorithm follows only one path at a time and therefore yields only one solution. For a given matrix, more than one solution could give the same minimum cost, or there could be a number of solutions with a minimal number of centers (but not necessarily minimum information loss). Furthermore, for a given set of centers, there could be a number of ways in which all the dialects could group with those centers. However, the optimization method always gives only one possible solution.

This drawback of the optimization method, that it gives only one solution, is countered in the least cost grouping algorithm of Section 3.3. The tradeoff is one of computing time. In the best case of one center, the least cost algorithm is much faster. In the worst case of *n* centers, the optimization algorithm is much faster than the least cost algorithm of Section 3.3.1, although with the refinements in Section 3.3.3 the latter would be faster than optimization. For the cases in between, it is not yet clear how they compare. I hope that the refinements suggested in Section 3.3.3 will make a computer program of the least cost algorithm run fast enough to be useful for large data matrices with complex solutions. If so, problems of degeneracy can be bypassed.

CHAPTER 4

EXPLAINING COMMUNICATION: A MODELING APPROACH

Chapters 4, 5, and 6 form a unit on the subject of explaining communication. The approach taken is one of building models. This chapter concentrates on the subject of modeling itself; the next two explore the two main components of a model for explaining communication: linguistic similarity and social relations.

Chapter 4 begins with a discussion of the meaning and advantages of modeling, especially with regard to explaining communication. This is followed by a consideration of the state of the art for the social sciences in general and for the communication problem specifically. Finally, a basic model for explaining communication is proposed. This model suggests that communication, or intelligibility, is based primarily on two factors: the linguistic similarity between dialects and the social relations between them.

In Chapter 5 the factor of linguistic similarity is considered in detail. Data from ten different field studies are analyzed in order to explore the relationship between lexical similarity and intelligibility. As a conclusion, a general model for expressing this relationship is proposed. Even though social relations are not incorporated, the model proves to be 70% accurate in predicting intelligibility from lexical similarity.

In Chapter 6 the second factor of the model, social relations, is considered in detail. After a general discussion, data from the island of Santa Cruz, Solomon Islands, are considered. A more comprehensive model which embraces social relationships as well as the linguistic ones of Chapter 5 is used to explain communication between dialects on the island. The predictions derived from this model are 90% accurate.

4.1 Why build models?

A model is a hypothesis about how something in the real world behaves. The models presented in the next two chapters are mathematical ones. This means that the hypotheses are stated in precise mathematical terms, in this case by numerical equations. Because the hypotheses are precise, they can be tested with precision. Herein lies the real value of the modeling approach: a hypothesis can be empirically tested against observed data with the result that the investigator knows exactly to what extent the model fits the data and to what extent it does not. When an acceptable model is found it can be used for one of two purposes: to explain the relations underlying what has already been observed, or to predict the value of a particular variable in the model when the values of the other variables are known. In the next two chapters, after the models for explaining communication are discussed, they are formulated mathematically, then tested empirically with field data. At the end we know exactly how much of communication is explained by the models and how much is not. But why should we want to build models to explain communication anyway?

Models for explaining communication can be applied to real world situations in at least three beneficial ways. First, they help us to understand patterns of communication. The techniques for measuring intelligibility discussed in Chapter 2 tell us only whether or not communication can take place and to what extent. The methods for analyzing patterns of communication discussed in Chapter 3 allow us to extract general patterns of communication and locate centers within the network of intelligibility relationships. However, neither method explains why there is communication at all or why the patterns of communication should be what they are. The models developed in the following chapters help us to do this. By understanding why patterns of communication are what they are, and not just what they are, leaders can make much better proposals about language planning in an area.

Secondly, the modeling approach, because it is also predictive, may shorten many of the logistic problems associated with intelligibility testing. An intelligibility survey is time consuming and sometimes difficult to carry out. If the level of intelligibility between dialects could be predicted, then we might be saved the task of trying to measure it.

Even when an intelligibility survey is carried out, it may not be feasible to test the intelligibility between all possible pairs of dialects when there are more than five or six dialects involved. In those cases, a predicting model can be used to estimate the untested intelligibility scores. For instance, in my survey of Santa Cruz Island (1977a), there were 13 dialects and measurements were made for 77 out of the possible 169 pairings, or 45% of the possible cases (Appendix 2.2, Table 2.2.6). In Kirk's Mazatec study in Mexico (Kirk 1970, Casad 1974:34), which involved 23 speech communities, intelligibility was tested for only 130 out of 529 possible pairings, or 25%. A predicting model can be used to estimate the intelligibility scores which are not actually tested.

This is of advantage not only for the sake of having a complete table of intelligibility relations to refer to, but is necessary if the analysis methods described in Chapter 3 are to consider all possible solutions. As noted already in Section 3.4, the method developed by Grimes (1974) has as one if its advantages that it does not require a complete matrix of values (1974:261). However, it has the disadvantage that when a value is missing, that particular reference point is excluded from serving as the center for that test point. This would have serious consequences in analyzing the results of the Santa Cruz survey, for instance, where intelligibility with the two central dialects (BAN and LWO) was seldom tested. There, intelligibility with the central dialects was a foregone conclusion based on the tests of more distant dialects. Unless values for the unmeasured intelligibility relations can be estimated, the analysis of communication centers may be skewed in the direction of those reference points most commonly tested for. Fortunately, a predicting model can be used to estimate the missing values and thus avoid this problem.

Finally, the predictive capability of a mathematical model may ultimately afford a more accurate estimate of intelligibility than intelligibility testing itself. A major unanswered problem with intelligibility testing is that of the adequacy of the text and the questions used for a particular test as a sample of the whole language. A short text can represent only an extremely small portion of the whole grammar and lexicon of the language. Even if all the problems associated with subject aptitude, subject screening, emotional reaction of the subject, and bilingual communication between the investigator and subject were completely absent or controlled for, there would still be no guarantee that the degree of intelligibility measured on the test was a good estimate of degree of understanding of the whole language. I feel that it is this point which requires the greatest faith in accepting intelligibility test results. If we understood the factors which underlie intelligibility well enough to construct a good predicting model, then that model could give predictions of intelligibility which were less skewed by the problems of subject and language sampling. Ultimately, predicting intelligibility may be more accurate than measuring it.

The modeling, or predicting, approach may not actually replace the intelligibility testing approach, at least not until we better understand the factors underlying intelligibility. For the present, the two approaches are complementary. Each serves as a check on the other. With predicted intelligibility (and informant opinions) serving as a backup to measured intelligibility and filling the gaps in it, less reliance on the measured intelligibility scores is required. Furthermore, the predicted scores can serve to point out measurement errors.

4.8 The state of the art

The development of modeling approaches in the social sciences is far behind its development in the physical sciences. In the physical sciences, a great many models have stood the tests of time and repeated confirmation, and have been elevated to the status of "laws" like Newton's laws of motion or Ohm's law. In the social sciences we are only beginning to use mathematical models to describe social phenomena.

John Q. Stewart, a proponent and developer of a field of study which he calls "social physics", traces the development of modeling in the physical sciences and shows its parallels in the social sciences. His social physics is an attempt to show that many sociological phenomena can be defined in terms of mathematical models, many of which are analogous to physical laws. He contrasts the current stage of development in the social and physical sciences as follows (1952:110):

> Merely verbal logic which traces back to Aristotle still comprises the sole intellectual equipment of too many practitioners of social disciplines, although physical science freed itself of those same archaic bonds as early as the seventeenth century.

Stewart traces the development in the physical sciences, and the parallels in the social sciences, in the following way (1947a:461):

> There was a time when scholars did not realize that number had the principal role in the description of the phenomena of physics. The transition from medieval to modern science was made in celestial mechanics, in three stages. These can be concisely represented by Tycho Brahe's extensive observations of planetary motions, Kepler's faith in mathematics as a means of insight into phenomena, and Newton's progress from Kepler's empirical rules for the solar system to the mechanics of the entire universe.
>
> We are now seeing a similar development in the social studies. Astonishing amounts of significant numerical data have been accumulated by conscientious social statisticians. Publications of the Bureau of the Census, for example, are comparable in extent and variety with catalogues of stars or tables of spectroscopic wave lengths, even if the numerical precision necessarily is much less. Thus the observational stage is well

advanced. A few investigators whose training is not confined to the social fields are beginning to proceed with the condensation of the voluminous sociological data into concise mathematical rules. The final rational interpretation of such empirical rules cannot come until after the rules themselves are established.

The three stages in the advance can be summarized as: (1) the collection of quantitative observations by Tycho Brahe, (2) their condensation into empirical mathematical regularities by Kepler, and (3) theoretical interpretation of the latter by Newton (from Stewart 1947b:179).

In the investigation of communication between speech groups, not even the first stage is well advanced. Quantitative observations on lexical cognate percentages between dialects all over the world are numerous, but quantitative observations on other aspects of linguistic relationships (such as phonology, grammar, and semantics) are scant. Quantitative observations of intelligibility between dialects are also rare, and observations on the social relations between dialects even more so. In Chapter 5 and Appendix 1 I gather all of the quantitative observations I could find in published and unpublished sources where both the percentage of intelligibility and the percentage of cognates are available. I could find such data from only ten language surveys around the world, a total of 245 observations. That is not enough data from which to derive universal laws, but it is enough to demonstrate that there are mathematical regularities in the relationship between intelligibility and lexical similarity.

The thrust of the next two chapters is along the second stage of development, namely, the condensation of observations into empirical mathematical regularities. I have encountered skeptics who feel that human relationships, such as communication between dialects, cannot be described in mathematical terms, because human behavior involves too many unknowns and irregularities. I trust that the empirical studies in Chapters 5 and 6 are sufficient to show that mathematical description is feasible, that the regularities are strong, and that the remaining unknowns play only a minor role. The third stage, that of interpreting the mathematical formulations and generalizing to universal laws, must wait until more observations from all over the world are available.

Before proceeding to present my own work in building models for explaining communication, I will report what others have done previously. In the first stage of model development, that of collecting quantitative observations, I am aware of only the following investigators who have reported quantitative observations on both intelligibility and linguistic similarity: Marvin Bender and Robert Cooper (1971) for Sidamo in Ethiopia, Bruce Biggs (1957) for Yuman in the United States, Eugene Casad (1974:78-81, 191-2) for Trique in Mexico, David Glasgow and Richard Loving (1964) for the Maprik area in Papua New Guinea, Warren Harbeck and Raymond Gordon (Harbeck ms [1969]) for Siouan in the United States and Canada, Peter Ladefoged (1968, Ladefoged and others 1972) for Bantu in Uganda, and Gillian Sankoff (1968, 1969) for Buang in Papua New Guinea. The data from all of the above studies, except from Glasgow and Loving, are reproduced in Appendix 1. Glasgow and Loving made impressionistic judgments of "mutual" intelligibility rather than acutally testing intelligibility in both directions. All these investigators report lexical cognate percentages as a measure of linguistic similarity. Bender and Cooper (1971) consider some grammatical relations as well, while Ladefoged (1968, 1970) quantifies phonological relations. Only three of the investigators -- Casad, Ladefoged, and Sankoff -- give any observations of relevant social factors; and only Casad (1974:191-2) quantifies these.

Only two investigators have entered into the second stage of model development, that of condensing the observations into mathematical regularities. Ladefoged computed the best fitting linear model for explaining his data and plotted it in a scattergram of the data points (Ladefoged and others 1972:76). Casad (1974:191-2) developed a linear model in three variables to explain intelligibility relations among five Trique dialects. The three variables are lexical similarity, intensity of contact, and location of contact. (The other two terms in his equation, the bilingualism factor and the error factor, are treated as constants.) The model fits the data very closely (97% explained variation, see Section 4.4). However, the model does not fit well with theoretical expectations. When there is no similarity and no contact, 6% intelligibility is predicted. Where there is 100% similarity and no contact 76% intelligibility is predicted. Where there is no similarity and complete contact, 46% is predicted. When there is complete similarity and complete contact, 121% is predicted. Our theoretical expectations for these four boundary conditions would be 0%, 100%, 100% and 100% respectively.

The work of Bender and Cooper (1971) should also be mentioned in this respect. Though they did not actually build models, they did explore regularities in the relationships between intelligibility, lexical similarity, grammatical similarity, and geographic proximity by computing correlation coefficients (see Section 4.4). The results showed that intelligibility correlated more highly with lexical similarity and geographic proximity than with grammatical similarity. Grammatical similarity was measured as the proportion of grammatical morphemes shared in translations of the same text (1971:42).

The third stage in model development, that of interpreting the results of stage two and generalizing to universal laws, has not been reached. There are publications, however, in which general models for explaining communication have been suggested. The models are not backed up by empirical validation and must therefore be viewed as exploratory.

The most elaborate of these is offered by Casad in an appendix to his book, Dialect Intelligibility Testing (1974:185-193). In his model, five independent variables underly intelligibility: (1) degree of linguistic similarity, (2) history of intragroup relations, (3) socioeconomic relations, (4) alternatives for language use, and (5) relative size of the groups. Five dependent variables intervene between the independent variables and intelligibility: (1) nature of intragroup contact, (2) societal attitudes, (3) language attitudes, (4) type of bilingualism, and (5) degree of bilingualism. The model is specified in terms of a directed graph which charts the cause and effect relations among the ten variables and intelligibility (1974:186). Twenty-six axiomatic propositions implied by the model are enumerated and sample theorems that can be derived from the axioms are given.

Ken Collier (1977) has proposed a simpler model. He suggests that intelligibility is an additive function of linguistic similarity and propensity to learn. The propensity to learn factor is a combination of two aspects of social relations, contact between dialects and the attitudes speakers of dialects have toward the other dialects. The paper includes suggestions on how the contact and attitude variables might be measured. Ronald Stolzfus (1974:43, 46) briefly suggests a similar model. He states that intelligibility results from the effect of linguistic similarity, or the effect of intergroup language learning, or the sum of both.

Two models which have been suggested for a closely related phenomenon, language change, are also relevant to the question of explaining communication. This is because the same kinds of variables which explain communication and language learning also seem to explain the borrowing aspect of language change. Again, these models are not backed up by empirical validation. The first was offered by Olmsted (1954b). His model predicts the likelihood that a single word will be understood and adopted by a single speaker. He suggests that the likelihood is an increasing function of the following factors: the degree to which the word is phonemically and morphemically regular in the hearer's system, the difference in social status of the speaker over the hearer, the upward social mobility of the hearer, the frequency of interaction between the speaker and hearer, and the frequency of occurrence of the word. He sums up the proposed model by saying that "the indispensables for lexical innovation are pronounceability and opportunity" (1954:115). In analogy to the models for explaining communication, these two indispensables are similarity and contact.

István Fodor (1965) has written a monograph entitled The Rate of Linguistic Change in which he develops a model for explaining language change. He discusses six factors involved in language change (pages 19-40): the historic effect, the cultural effect, the social effect, the geographic effect, the effect of neighboring and distant foreign peoples, and the role of national character. In addition he discusses possible ways of measuring language change by quantitative methods (pages 41-58) and a mathematical model of the rate of linguistic change (pages 59-73).

4.3 A basic model for explaining communication

Everyone who has tried to explain communication agrees on at least one thing, that two main factors play a key role in determining the presence or absence of communication: language variation and the social setting. On the one hand, the degree of intelligibility between two dialects is related to linguistic similarity. The greater the similarity, the greater the intelligibility is likely to be; conversely, the lower the similarity, the lower the intelligibility is likely to be. On the other hand, the degree of intelligibility is related to the social setting in which the communication occurs. If the social situation is favorable, contact and learning will lead to a boost in intelligibility. If the social situation is not favorable, it will tend to limit intelligibility. Thus intelligibility can be viewed as comprised of two components: a linguistic, or similarity-based, component and a social, or contact-based, component. That is,

total intelligibility =
similarity-based intelligibility +
contact-based intelligibility

This formulation with a simple addition is oversimplified; however, it serves as a useful starting point for discussion.

It is on the specifics of what factors go into each of the two components of intelligibility, how these factors can be measured, and how the components interact, that investigators have differed. The discussions of the subject mentioned in Section 4.2 have been largely exploratory and based on little supporting evidence. The next two chapters consider both of these components and demonstrate with empirical evidence how they can be measured and built into models for explaining communication.

4.4 Some statistical preliminaries

Formulating and testing mathematical models involves the use of statistics. In this section, the basic statistics referred to in the next two chapters are briefly defined. For a complete discussion of these statistics and how they are computed, the reader is referred to a basic text on statistics, such as Blalock 1972, Darlington 1975, or Downie and Heath 1974.

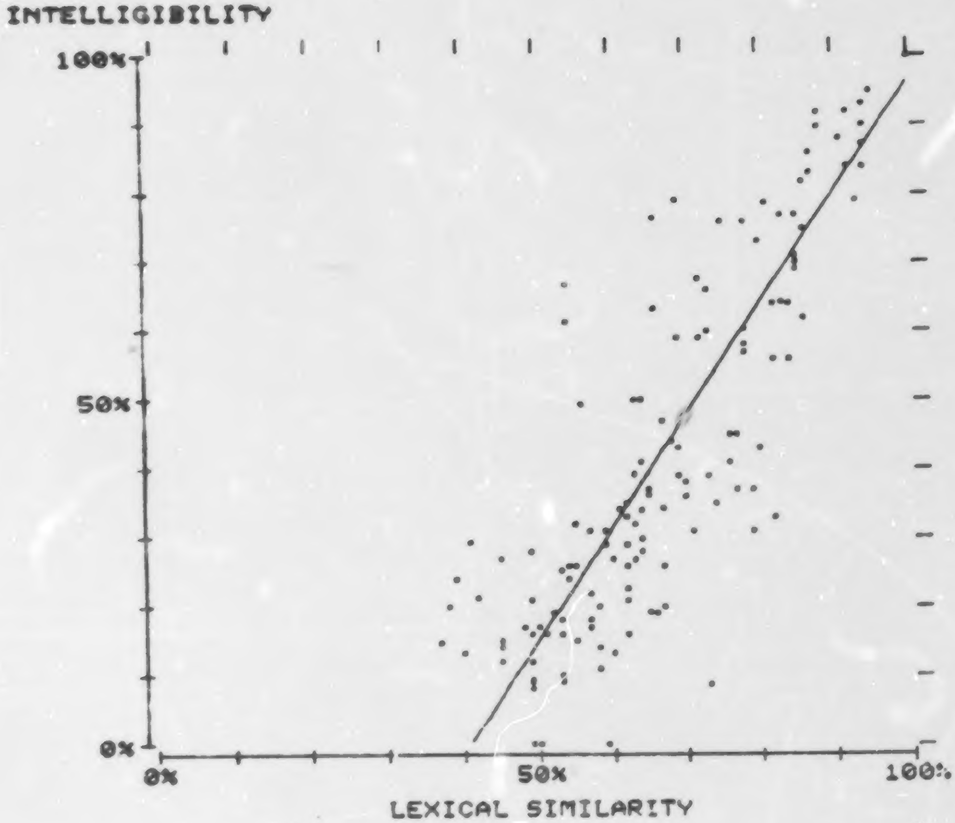
The standard statistical techniques of least-squares regression and correlation form the basis of the analysis in Chapter 5. In a nutshell, these techniques are used to test how well the values of one variable predict the values of another. The variable being predicted is called the *dependent variable*, and a variable used as a predictor (there may be more than one) is called an *independent variable*. The techniques are easiest to visualize if the data are plotted in a two dimensional graph. Figure 4.1 gives an example (it is copied from Figure 5.8 in Section 5.2.6). In Figure 4.1 the percentage of intelligibility is plotted on the vertical axis while the percentage of lexical similarity is plotted on the horizontal axis. Each case in the data consists of a pair of observed values, the intelligibility from one dialect to another and the percentage of cognates they share. In the graph a dot is placed where the paired values of intelligibility and lexical similarity intersect. The plotted points are scattered within the graph and for this reason such a graph is called a *scattergram*. Note, however, that the scattering is not random; there is a pattern.

Regression analysis is used to fit a curve to that pattern. When performing regression analysis, one first selects the desired shape of the curve; that is, whether it will be a straight line, a parabola, an exponential growth curve, and so on. The analysis then determines the parameters for the curve of that shape which most closely describes the data. Most of the regressions performed in Chapter 5 are linear. Linear regression finds a single straight line which best fits the pattern of the scattered points. In doing so, it finds the straight line passing through the data points in such a way that the average square of the distance of the data points from that line is the least possible. This line is called the regression line. Figure 4.1 illustrates the best fitting linear regression line for the given data points. It can also be thought of as a prediction line; the predicted value of the dependent variable can be read from the intersection of the regression line with the given value of the independent variable.

Correlation analysis measures the amount of scatter about the regression line. It is therefore used to assess the goodness of fit of the line and the model it represents. The correlation coefficient used in this analysis is the Pearson product-moment correlation coefficient, symbolized with $\underline{r}$. The absolute value of this coefficient ranges from zero for no correlation to one when all of the data points lie exactly on a straight line. Thus, when the points cluster close to the regression line, the correlation coefficient approaches one. When the points are scattered far from the line, the correlation coefficient approaches zero.

When the correlation coefficient is less than one, it is an indication that predictions of the dependent variable made with the regression line are not perfect. The *standard error of estimate* measures the amount of prediction error associated with the predictions. It is used to compute an interval estimate for predictions. For instance, to say that predicted intelligibility is 80% is to use a point estimate; to say that it is between 70% and 90% is to use an interval estimate. When the standard error of estimate is used to compute an interval estimate, the interval is characterized by a *confidence level*. The standard error of estimate

Figure 4.1 Scattergram and regression line



itself defines a 68% confidence interval. This means that in 68% of the cases the true value of the dependent variable is within a range of plus or minus one standard error of estimate from the predicted value. Doubling the standard error of estimate defines a 95% confidence interval within which the true value can be expected to lie 19 times out of 20. For instance, if the predicted value of intelligibility is 70% and the standard error of estimate is 8%, then we can say that the true value of intelligibility is within the range of 54% to 86% with 95% confidence. The multiplicative constants for defining other levels of confidence can be found by consulting any statistics text book.

The significance of the correlation coefficient offers a means of evaluating the degree of confidence in the strength of the relationship between two variables. It tells us how much trust we can put in the correlation coefficient and the regression line. It is possible that two variables could be totally unrelated but that the chance distribution of the two randomly related variables would yield a high correlation coefficient. As the number of data points increases, the likelihood of a spurious correlation decreases. The significance of the correlation coefficient is computed as a probability. It is the probability that the value of a correlation coefficient as large or larger than the one calculated could have arisen by chance alone, were the two variables in fact uncorrelated. For instance, a significance level of .001 means there is a one in a thousand chance that the observed relationship between the variables could be due to chance alone. In the social sciences, a significance level of .05 or less is generally considered to be significant. A significance level of .05 is the same as a confidence level of 95%.

A final statistic for evaluating the strength of relationship between two variables is the percentage of explained variation. In the data of the next chapter, measured intelligibilty varies from 0% to 100%. At the same time lexical similarity varies from 0% to 100%. In doing the statistical tests described above, we are asking, "Can the variation in measured intelligibilty be explained by the variation in lexical similarity?" That is, when lexical similarity goes up, does intelligibity also go up, and by a proportional amount? By the same token, when lexical similarity goes down, does intelligibility also go down, and by a proportional amount? The percentage of explained variation answers these questions directly. The percentage of explained variation tells how much of the measured variation in intelligibility is explained by the variation in lexical similarity, or, what percentage of the ups and downs in intelligibility correspond to ups and downs in lexical similarity.

In evaluating the adequacy of explaining (or predicting) models, the total variation in the dependent (or predicted) variable is partitioned into two components, the explained variation and the unexplained variation. That is,

total variation =
explained variation +
unexplained variation

In the statistical analyses which follow, the total variation in the dependent variable is measured by its sum of squares -- the sum of the squared differences between the actual values of the dependent variable and its mean value. (When the sum of squares is divided by the number of cases, the result is a statistic called the variance. Thus the percentage of explained variation I am using is equivalent to the percentage of explained variance.) The explained variation is measured by the regression sum of squares -- the sum of the squared differences between the predicted values and the mean value. The unexplained variation is measured by the residual sum of squares -- the sum of the squared differences between the predicted values and their corresponding actual values.

The percentage of explained variation is computed by dividing the explained variation by the total variation and multiplying the result by 100. When the correlation coefficient is squared, the result is the proportion of explained variation. Thus another way to compute the percentage of explained variation is to square the correlation coefficient and multiply by 100. The percentage of unexplained variation can be computed by subtracting the percentage of explained variation from 100%.

For the problem of explaining intelligibility as a function of lexical similarity, the partitioning of variation is as follows:

total variation in intelligibility =
variation explained by lexical similarity +
unexplained variation

If there were no unexplained variation, then the model would be complete. Variation in lexical similarity would explain all of the variation in intelligibility, and we would say that lexical similarity is a perfect predictor of intelligibility. However, when the percentage of explained variation is less than 100% then lexical similarity is not a perfect predictor of intelligibility and the model is incomplete. A model is complete only if it can account for all the total variation. To complete the model, we must introduce additional factors to explain the unexplained variation. If the unexplained variation is small it can be attributed to measurement error, either in test construction and scoring, in sampling, or in both. When the unexplained variation is greater, however, measurement error alone can no longer be used to account for the unexplained variation. At this point it is necessary to introduce other factors into the predicting model, such as social factors or other aspects of linguistic similarity, or to change the mathematical relations in the model, such as from linear to exponential.

In the next two chapters, the attempt is made to explain communication. The approach is one of successive refinements. In each chapter a succession of models is considered. At each step refinements are made by incorporating new or different factors or different mathematical relations into the model in order to account for a portion of the previously unexplained variation, and thus increase the percentage of explained variation.

CHAPTER 5

EXPLAINING COMMUNICATION: LINGUISTIC FACTORS

This chapter considers how the linguistic similarity between dialects affects the intelligibility between them. In Section 5.1 the discussion covers the general problems of quantifying linguistic similarity so that it can be incorporated into a mathematical model. In Section 5.2, an empirical analysis of the relation between lexical similarity and intelligibility is made. This analysis is based on data gathered in ten different field studies throughout the world. As a final conclusion, the possible universal relationship between lexical similarity and intelligibility suggested by the concurring sets of field data is explored.

5.1 Quantifying linguistic similarity

The approach of modeling by numerical equation requires that we describe linguistic similarity numerically. However, linguistic similarity is not an easy concept to quantify. Languages may differ in their sound systems, their vocabularies, their grammars, or their semantic systems. Because linguistic similarity is such a complex relationship, it is impossible to summarize it completely in one number, at least at the present time. This is one of the motives behind the early studies of intelligibility. They hoped by testing intelligibility to discover a means of indirectly quantifying linguistic similarity, or "dialect distance" as they called it (Pierce 1952, Biggs 1957). However, their perspective was backwards (Wolff 1959). Intelligibility does not determine linguistic similarity; rather, linguistic similarity along with other factors determines intelligibility. Thus the burden falls back on finding a means to quantify linguistic similarity directly.

Many techniques have been proposed for quantifying specific aspects of lingusitic similarity. The most widely used is lexicostatistics, which measures the degree of similarity in basic vocabulary between languages. The method was developed by Morris Swadesh (1950, 1952, 1955, also Lees 1953). Helpful discussions are given by Gleason (1959), Gudschinsky (1956), Hymes (1960), and Sanders (A. Sanders 1977a).

A number of methods for quantifying phonological similarity, or phonostatistics, have been proposed. However, none has gained the widespread use and acceptance that lexicostatistics has. This is probably because the development of phonostatistics was nearly ten years later and because phonostatistics is computationally more complex. The most promising methods have been developed by Grimes and Agard (1959), McKaughan (1968), and Ladefoged (1970; Ladefoged and others 1972:62-65). Elsewhere I give a review of these and nine other phonostatistic methods (Simons 1977c).

A few attempts at quantifying grammatical similarity have been made but with limited success. Again, these methods have not enjoyed a widespread use or acceptance. In general, these grammatical methods require a good analysis and understanding of the grammars which are being compared. For this reason they are not applicable to the language survey situation, unless the investigator has a very good idea of what the grammars will be like on the basis of comparative study. Methods of grammatical statistics have fallen into two major categories. The first computes measures of association between dialects by comparing them for the presence or absence of key morphological or syntactic features (Kroeber and Chrétien 1937, 1939, Ellegård 1959, Simons 1977c:172-3, see also Capell 1962). The second computes typological indices which characterize single dialects as to their position along some dimension of language structuring. For instance, an "index of synthesis" measures the average number of morphemes per word. Comparisons between dialects are achieved by comparing their indices (Greenberg 1960, Kroeber 1960, Voegelin and others 1960, Voegelin 1961, Moore 1961). Bender and Cooper (1971) used a third method which resembles lexicostatistics more than either of the above typological methods. Their intelligibility tests were based on six texts that were translated into each of the six dialects they were testing (see Section 2.2.4). They were thus able to make morpheme by morpheme comparisons of the translated texts and compute the percentage of grammatical morphemes (as opposed to root morphemes) which were the same for each pair of dialects. These measures of grammatical association were then correlated with measured intelligibility; the results were largely inconclusive.

Quantifications of semantic similarity have not yet been used by linguists to my knowledge. Such a method could follow the first method described above for grammatical statistics. Each pair of dialects would be compared for the presence or absence of key semantic oppositions. The work of Berlin and Kay (1969) on color terms contains the information and analysis necessary to quantitatively compare 98 languages of the world on the semantics of their color terminology. Furthermore, their work develops a methodology which could be applied for the remaining languages. Other semantic domains which have been well studied are kinship terminology and body part terminology. Another possible approach is Charles Osgood's semantic differential technique, which is a method for quantifying and comparing meaning (Osgood and others 1957, Snider and Osgood 1969).

At the present time, the prospects for a composite quantification of linguistic similarity are not good. A number of phonostatistic methods exist, but none has been widely used, mainly because the computations are complex. Good techniques for gathering and quantifying data on grammatical and semantic similarity, at least in the dialect survey situation, are still in the future.

Lexicostatistics remains as the most widespread and readily available means for quantifying linguistic similarity. The analysis in the next section of this chapter, especially Section 5.2.5, demonstrates that lexical similarity is a good predictor of intelligibility and thus must be viewed as a useful approximation to a measure of linguistic similarity. Nevertheless, many investigators have avoided or belittled the use of lexicostatistics. There are at least three reasons for this.

First, the pitfalls of glottochronology with its assumptions of a universal rate of change and the requirement of independent change, and the ensuing misuse of lexicostatistics in studies of linguistic history, have tainted the image of lexicostatistics. However, if we take lexicostatistics at face value for merely what it is, "word statistics", it is free from these assumptions and problems. Under these conditions it actually proves to be an effective predictor of intelligibility. That is, similarity of basic vocabulary is a more reliable indicator of intelligibility between languages than it is of the historical time depth between languages. Elsewhere (Simons 1977d:14-17) I have contrasted the methods of synchronic lexicostatistics and diachronic lexicostatistics and shown that the future of the synchronic use of it is bright while that of the diachronic use is not.

Second, lexical similarity is only one aspect of linguistic similarity. Some investigators have thus been leery of depending on it to estimate linguistic similarity. However, the results in the next section indicate that lexical similarity alone is a good predictor of intelligibility, and therefore approximates linguistic similarity as well. The results do not suggest that phonological, grammatical, and semantic similarity are not important, but simply that degree of lexical similarity parallels the degree of phonological, grammatical, and semantic similarity. This would imply that change in these other aspects of language tends to keep abreast of change in vocabulary. This is not always the case, but it probably averages out. For instance, Grimes (1974:267) has shown that French and Catalan group more closely with Spanish and Portuguese than with Italian on the basis of phonostatistics (Grimes and Agard 1959, Grimes 1964) but they group more closely with Italian on the basis of lexicostatistics (Rea 1958). The reason is that the one measure is sensitive to a heavy lexical borrowing in French from Italian around the Renaissance period, while the other measures sound change. However, for the rest of Romance, the two groupings agree.

Finally, it has been suggested that lexicostatistic measures are not as appropriate as phonostatistic measures in assessing linguistic similarity for degrees of language divergence where intelligibility is still expected. McKaughan (1964), in an analysis of linguistic relations among a number of dialects in the New Guinea highlands, used three methods: lexicostatistics, phonostatistics, and structural comparison. In conclusion he suggested that each method was most useful within certain ranges of linguistic divergence: phonostatisic methods are most applicable where there is slight divergence, lexicostatistic methods where there is moderate divergence, and structural comparisons where there is wide divergence (McKaughan 1974:118). Ladefoged (1968:5, Casad 1974:118-9) has suggested that since we expect intelligibility only between highly similar dialects, phonostatistic methods may be more useful than lexicostatistic or grammatical methods in predicting intelligibility. On the basis of these suggestions and possibly the other two factors mentioned already, Casad (1974:118-119) does not even consider lexicostatistics in his chapter on alternative approaches for assessing intelligibilty.

The results in Section 5.2 do not prove or disprove McKaughan's hypothesis. They do show, however, that any assumption that lexicostatistic measures are not sensitive enough within the range of linguistic divergence apropriate to the range of intelligibility is ill founded.

5.2 Lexical similarity and intelligibility

5.2.1 Overview of the data and method

This study of the relationship between lexical similarity and intelligibility is based on ten field studies conducted in various parts of the world. These studies were conducted by ten investigators in ten different language groups. The groups span three continents -- Africa, Oceania, and North America. The specific areas involved are Ethiopia, Uganda, Papua New Guinea, the Polynesian islands, Mexico, Canada, and the United States. Not only were the circumstances of each of the studies different; so were the methodologies. In spite of all these differences, the degree of convergence between the results of all these field studies is very striking. Section 5.2.4 shows that eight of these studies point to almost exactly the same underlying relationship between lexical similarity and intelligibility.

In each field study the percentage of intelligibility between dialects in the study area was measured. Corresponding to each measurement of intelligibility is a measure of the lexical similarity between the same two dialects. These measures are expressed as a cognate percentage. Each pair of measurements, an intelligibility percentage with a cognate percentage, is treated as one case in the statistical analysis. The smallest study contains only nine cases, while the largest study contains seventy-seven. The average size is twenty-four cases. The complete details about each study, including the sources of the data, some notes on the methodologies used, and a listing of the raw data are found in Appendix 1.

The analysis begins in Section 5.2.2 by examining the results obtained from the raw data. In Section 5.2.3 the analysis is refined by removing some of the effects of social factors from the predicting model. In Section 5.2.4 the prediction of intelligibility by lexical similarity is further sharpened by adjusting the intelligibility scores for measurement error. Final conclusions are drawn in Section 5.2.5. In Section 5.2.6 the data from the different field studies are pooled and possible models for the universal relationship between intelligibility and lexical similarity are explored.

Except for Section 5.2.6, the method of linear regression is used throughout the analysis to find the relationship between similarity and intelligibility. This makes the assumption that the relationship between the two variables is a linear, or straight line, one. A straight line plot says that a given amount of increase in lexical similarity will give the same increase in intelligibility at any point along the intelligibility scale. There is no theoretical reason why we should expect this to be the actual case. For instance, the factor of redundancy would suggest that an increase in similarity would have less and less of an effect on intelligibility as the intelligibility neared 100%. However, the scattergrams in Appendix 1.3 show no consistent hint of nonlinearity. Thus linear techniques were used in the analysis since they are computationally the simplest. The assumption of linearity is a weaker one than the assumption of nonlinearity and is thus appropriate for a first approximation. The use of nonlinear techniques should increase, not decrease, the degree of fit of the models. In Section 5.2.6, the data from eight studies are pooled and nonlinear relationships are explored. Nonlinear models turn out to offer a slight, but not statistically significant, improvement over the linear model for the current data.

5.2.2 Results from the raw data

The data have been briefly described already in Section 5.2.1. In Appendix 1.1, each of the ten sets of data is described more fully. In Appendix 1.2, all of the data is listed. In Appendix 1.3, a scattergram showing the distribution of intelligibility versus lexical similarity is plotted for each of the ten studies. Below each scattergram, the following figures (see Section 4.4) are listed: the number of cases, the correlation coefficient, the significance, the standard error of estimate, and the percentage of explained variation. In addition, the line of best fit given by the regression analysis on the full data is drawn into the scattergram as a solid line. The formula for this line is given at the base of the scattergram.

From the formula for the regression line, it is possible to compute two other helpful quantities. The first is the predicted value of intelligibility when lexical similarity is 100%; the second is the value of lexical similarity when the predicted value of intelligibility is 0%. The first quantity, the predicted value of intelligibility for 100% lexical similarity, gives a measure of naturalness for the prediction equation. The regression line should predict 100% intelligibility when lexical similarity is also 100%. The nearness of the predicted value to 100% gives a measure of naturalness for the prediction equation. The second quantity, the value of lexical similarity for a predicted intelligibility of 0%, offers a means of comparing the convergence of the ten different studies. Ideally, at the upper end of the regression line, the lines for all ten studies should converge on the point at (100%, 100%). At the lower end, however, where the lines intersect the similarity axis, the lines fan out indicating the differences between studies. The points at which the predicting lines intersect the similarity axis give a good means of comparing the degree to which the regression lines from the different studies are the same or different.

In Figure 5.1 the regression lines from the ten different scattergrams in Appendix 1.3 are superimposed on the same graph. Note that all ten studies show the same general trend, a regression line which starts in the lower left and rises to the upper right. There is a general convergence toward the (100%, 100%) point; however, it is not very strong. The predicted values of intelligibility for 100% lexical similarity range from 68% to 102%. This explains most of the crisscrossing of the prediction lines.

In Figure 5.2 the key statistics from Appendix 1.3 are compiled into a summary table. For each of the ten studies, the following figures are given: the number of cases (N), the percentage of explained variation (%EV), the correlation coefficient (Corr), the significance of the correlation coefficient (Sig), the standard error of estimate (SEE), the predicted intelligibility for 100% lexical similarity (Lex-100), and the lexical similarity for 0% predicted intelligibility (Int-0). In the top portion of the table the figures for each of the ten studies are given; in the bottom portion they are summarized. Four figures are given in the summary: the minimum observed value, the maximum observed value, the mean (or the average) of the ten observed values, and the standard deviation from the mean. The standard deviation is a measure of dispersal around the mean. Roughly speaking, it tells the average amount by which the observed values differ from the mean.

The data can be summarized as follows. The ten studies contain, on average, 24 cases. The percentage of explained variation ranges from 18% to 97% with an average of 65%. The average correlation coefficient is .79140. In only one study, Biliau, is the significance doubtful; in all other cases the probability of a spurious correlation is less than one in a thousand. The average standard error of estimate for predictions of intelligibility is 13%. The predicted values of intelligibility for 100% lexical similarity range from 69% to 102%, with an average of 90%. The standard deviation of 84% for the points at which the regression lines cross the similarity axis, gives an indication of how scattered the prediction lines based on raw data are.

5.2.3 Controlling for nonsymmetric social factors

In the average case, lexical similarity alone accounts for 65% of the variation in raw intelligibility scores. 35% of the variation in intelligibility remains unexplained. In this section, almost one half of this unexplained variation is

Figure 5.1 Plots for full raw data

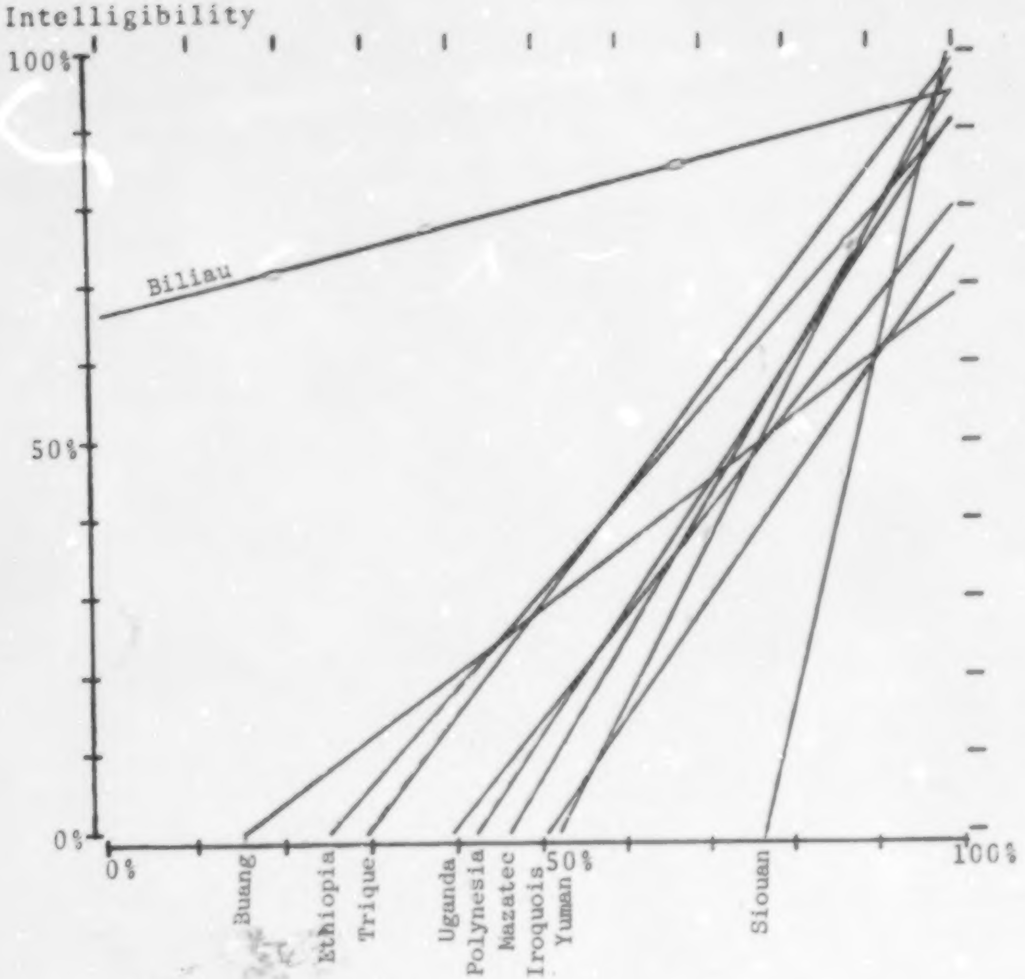


Figure 5.2 Statistics for full raw data

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
Biliau	9	18.1	.42487	.2543	6.1	94.7	-233.5
Buang	21	49.3	.70232	.0004	11.8	68.8	15.3
Ethiopia	30	71.6	.84592	.0001	16.2	91.2	25.1
Iroquois	14	66.0	.81267	.0004	21.0	75.0	50.6
Mazatec	19	65.1	.80659	.0001	13.1	95.1	46.1
Polynesia	77	74.6	.86350	.0001	14.4	91.6	42.3
Siouan	25	64.9	.80543	.0001	18.1	102.5	76.6
Trique	15	58.5	.76503	.0009	11.2	99.2	29.4
Uganda	10	81.8	.90457	.0003	12.8	80.3	39.4
Yuman	25	96.6	.98310	.0001	7.0	97.9	52.0
Minimum	9	18.1	.42487	.0001	6.1	68.8	-233.5
Maximum	77	96.6	.98310	.2543	21.0	102.5	76.6
Mean	24	64.6	.79140	.0257	13.2	89.6	14.3
Deviation	19	19.8	.14198	.0762	4.4	10.6	84.1

attributed to social factors, specifically, to the effects of nonsymmetric social relations which can be observed in the intelligibility data.

A basic model for explaining intelligiblity has already been introduced in Section 4.3. There it was suggested that intelligibility has two components, a linguistic similarity-based component and a social contact-based component. In terms of a partitioning of variation this model can be expressed as,

total variation in intelligibility =
variation explained by linguistic factors +
variation explained by social factors

In the previous section, we investigated only the contribution of linguistic similarity (specifically, lexical similarity) to explaining intelligibility. It then follows from the preceding formula that the variation due to social factors is as yet a component of the unexplained variation.

The data do not include measurements of relevant social factors; therefore, it is not possible to do a full investigation of the contribution of social factors. However, there is one property of intelligibility which points to the presence of social factors and that is nonsymmetry. Dialect A may understand B better than B understands A, or vice versa. According to our basic model this must be explained by the presence of nonsymmetric relations of linguistic similarity or nonsymmetric social relations. Lexical similarity, our current approximation to linguistic similarity, is a symmetric measure. That is, the percentage of cognates from B to A is always the same as that from A to B. If there are any nonsymmetric linguistic factors these also would appear in the model in the unexplained category.

There are therefore two possible hypotheses: that nonsymmetric intelligibility relations are explained by nonsymmetric linguistic relations or by nonsymmetric social relations. I am assuming in these data that they are due to nonsymmetric social relations. The sources do provide some evidence for this, while they provide no evidence for the alternative hypothesis that nonsymmetric intelligibility is explained by nonsymmetric linguistic relations. Of the ten studies, only Sankoff address the latter possibility but concludes that there is no basis for accepting the hypothesis. She observes that ffor the Buang data, explaining "non-reciprocal intelligibility ... on the basis of phonetic differences between the codes gives equivocal results" (1969:847, 1968:183). Other writers, for instance Wurm and Laycock (1961:129-132) and St. Clair (1974a:93-5,1974b:146-7), have attempted to explain nonreciprocal intelligibility in terms of asymmetric linguistic relations, but their evidence is impressionistic rather than empirical. While I do not deny that linguistic relations contribute to nonreciprocal intelligibility, the evidence which demonstrates the extent to which they do is presently lacking.

The sources do give evidence for nonsymmetric intelligibility caused by nonsymmetric social relations. For the Biliau data, which were collected by my wife and myself, intelligibility relations in the direction of Biliau village are greater than those directed away from Biliau. This is because that village is the political and economic center for the region. At Biliau are located an airstrip, a harbor, a primary school, a medical clinic, and a mission station. For the Buang data, Sankoff (1969:847) notes that the nonsymmetric intelligibility is explained by contact arising from travel routes down the river valley toward the government station. For the Ugandan data, Ladefoged, Glick, and Criper (1972:76) observe that in the one case of nonsymmetric intelligibility, the better understood dialect "is spoken in the capital of the country, and has more time on the radio than any other

Ugandan language."

The presence of these nonsymmetric social relations shows up in the intelligibility relations as significantly different scores for communication in both directions between the same two dialects. In such cases we assume that the higher intellligibility score of the pair is boosted by nonsymmetric social relations (that is, boosted by contact and learning). By removing cases where this boosting is detected, it is possible to control for the contribution that nonsymmetric social relations make to explaining intelligibility.

Nevertheless, there still remain cases where a social relation that is symmetric can boost intelligibility in both directions and go undetected by this method. A good example is the Biliau data. In that study the two most divergent dialects are only three hours' walking distance away and there is a lot of contact in both directions. These cases must be relegated to the category of unexplained variation.

A more complete model for the decomposition of variation in intelligibility is now,

> total variation in intelligibility =
> variation explained by lexical similarity +
> variation explained by nonsymmetric factors +
> unexplained variation

where unexplained variation includes nonlexical aspects of linguistic similarity, symmetric social relations, and measurement error.

This model suggests that if the effects of nonsymmetric social factors can be controlled, then unexplained variation will decrease. This hypothesis can be tested with the data from the ten field studies. The method used is to remove cases from the sample in which a boosting of intelligibility due to nonsymmetric social factors is suspected, and then to repeat the correlation and regression analysis. Such cases were found by inspecting the data. First the symmetric pairs of cases were found. A symmetric pair of cases is two cases which measure communication in both directions between the same two dialects. If one of the intelligibility scores in the symmetric pair is significantly higher than the other, then that case is dropped from the sample. To judge a significant difference, it was not possible to make tests of significance since the reported data do not contain standard deviations for the intelligibility measurements. Instead a simple rule of thumb was used: if one score was 10% or more greater than the other then it was considered to be significantly higher. The cases thus removed from the sample are indicated in Appendix 1.2 by an "X" in the "Excluded" column. In the scattergrams in Appendix 1.3, the excluded points are plotted as "X" while the remaining points are plotted as circles. Examination of the scattergrams shows that the excluded points, in general, lie well above the regression line. There are points, however, which are further from the regression line than the excluded points. These are probably examples of undetected symmetric social factors which boost intelligibility.

In the scattergrams in Appendix 1.3, a second regression line is drawn in as dashed line. This is the regression line for only those points plotted as circles. Below the scattergrams two sets of statistical computations are given. The first set is for all of the data points; the second set is for the circle points only, the points which remain when the "X" points are excluded. The statistics computed for the data with exclusions are compiled into a summary table in Figure 5.4. The format of this table is the same as that of Figure 5.2 explained previously. In Figure 5.3, regression lines for the ten studies with the "X" points excluded are superimposed in one graph. This graph parallels Figure 5.1.

The effect on the results of controlling for nonsymmetric social factors can be seen by comparing Figure 5.4 with Figure 5.2. The average number of cases is reduced from 24 to 20; thus, on average, four cases were removed from each study. The change in percentage of explained variation is substantial; it rises from 65% to 81%. This 16% additional explained variation supports the original hypothesis that nonsymmetric social factors are an important element in explaining intelligibility. The other measures of predicting accuracy and reliability show comparable improvements: the correlation coefficients increase by .10564 on average, the average significance improves nearly ten times, and the standard error of estimate decrease from 13% to 10%. There is no significant change in the average predicted value of intelligibility for complete lexical similarity; it is still below 90% with a standard deviation exceeding 10%. There is, however, an improvement in the degree to which the prediction lines of the different studies fan out; this is seen by the decrease from 84 to 40 in the standard deviation of the point at which the lines cross the similarity axis.

5.2.4 Controlling for intelligibility measurement error

Not all aspects of measurement error need be classified as unexplained variation. One aspect of measurement error in intelligibility scores can be adjusted for. In the administration of intelligibility tests, subjects seldom get a perfect result on the test from their own dialect. However, they theoretically should understand their own form of speech perfectly. When test results indicate that they do not, these results are best interpreted as pointing to deficiencies in the abilities of the subject, in the construction of the test, or in the administration of the test. It is possible to control for these kinds of measurement errors by adjusting raw intelligibility scores on the basis of performance on the hometown test. The kinds of measurement errors which still lie beyond the reach of such adjustments are sampling errors; that is, those which have to do with how well the group of subjects represents the whole community and how well the text represents the language as a whole.

The need for hometown score adjustments in the data of the ten field studies is seen in Figure 5.5. This table shows the distribution of hometown scores for each of the ten field studies. The first column gives the lowest measured hometown score, the second column gives the highest hometown score, and the third column lists the average hometown score. Note that in the case of the Buang study the average hometown score is only 69%. Taken at face value, this suggests that the Buang villager can understand only 69% of what his neighbor says to him. This obviously is not true. On the other hand, in the Trique study the average hometown score is as high as 98%. The last row of Figure 5.5 shows that overall the hometown scores range from a low of 46% to a high of 100% with an average hometown score of 90%. The wide difference in average hometown scores between individual studies accounts for the scattering of the regression lines in Figures 5.1 and 5.3 in the top right hand corner of the graphs. Theoretically, all the lines should converge on the point (100%, 100%). However, because the average hometown score varies from 69% to 98%, so do the predicted values of intelligibility when lexical similarity is 100%.

Adjusting the raw intelligibility scores in such a way that hometown scores are

Figure 5.3 Plots for raw intelligibility with exclusions

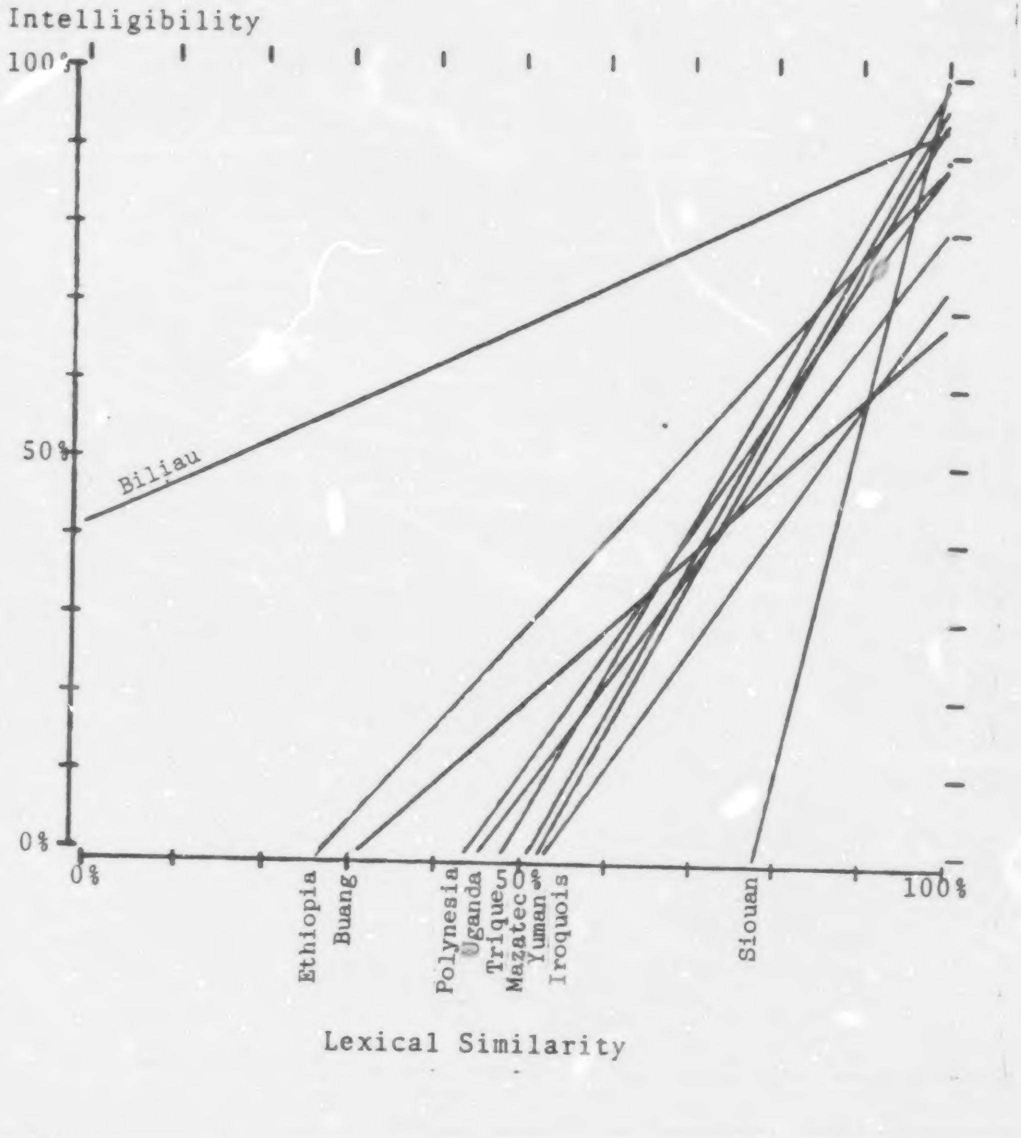


Figure 5.4 Statistics for raw intelligibility with exclusions

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
Biliau	6	74.2	.86156	.0274	3.2	93.4	-79.9
Buang	15	65.8	.81090	.0002	10.7	68.0	31.1
Ethiopia	23	75.1	.86677	.0001	16.1	88.5	26.8
Iroquois	12	80.9	.89944	.0001	15.8	72.7	52.8
Mazatec	17	71.7	.84672	.0001	12.1	96.3	50.8
Polynesia	67	83.0	.91091	.0001	11.5	88.3	43.5
Siouan	20	74.2	.86156	.0001	15.9	100.6	77.9
Trique	11	88.7	.94174	.0001	6.7	98.9	47.8
Uganda	9	96.1	.98010	.0001	6.3	80.1	45.1
Yuman	21	98.1	.99066	.0001	5.2	94.4	52.2
Minimum	6	65.8	.81090	.0001	3.2	68.0	-79.9
Maximum	67	98.1	.99066	.0274	16.1	100.6	77.9
Mean	20	80.8	.89704	.0028	10.4	88.1	34.8
Deviation	16	10.1	.05590	.0082	4.5	10.6	40.4

Figure 5.5 Distribution of hometown scores

	Lowest	Highest	Average
Biliau	92	95	94
Buang	67	73	69
Ethiopia	81	99	91
Iroquois	46	83	73
Mazatec	89	100	94
Polynesia	93	98	96
Siouan	87	100	93
Trique	97	99	98
Uganda	79	82	80
Yuman	84	96	92
Combined	46	100	90

raised to 100% has two effects. First, by compensating for measurement error it decreases the amount of unexplained variation in the model. Second, and most important, it makes the results of different field studies more comparable. When the hometown scores in the Buang study average 69%, while the hometown scores in the Trique study average 98%, it is very difficult to compare the two studies to determine if they suggest a common trend. However, when all of the intelligibility scores are adjusted to raise the hometown scores to 100%, the results of all the different studies are put on the same scale of measurement. They can then be compared directly to one another and the cases from the different studies can even be joined into one large set of data. The effect in the plot of regression lines (shown in Figure 5.6) is that the lines converge much more sharply toward the (100%, 100%) point.

The discrepancy between the hometown score and 100% can be attributed to one of three things: a learning curve, the subject's abilities, or test deficiencies. Depending upon the source of the discrepancy, three different methods can be used to adjust the intelligibility scores to normalize the hometown scores to 100%. The three sources of discrepancy and the methods used to compensate for them are as follows:

(1) Learning curve - The hometown test should be the first test which a subject takes. This is so he can learn to take the test without having to contend with dialect differences at the same time. In spite of efforts to explain how the testing will be done and of even having a preliminary warm-up test, it could be the case that the subject was still learning how to take the test when he took the hometown test. This could result in errors on the hometown test. We may be able to assume that these errors affect only the hometown test and by the time the subject gets to the second test there will be no more such errors. The solution for adjusting intelligibility scores in this case is to raise all the hometown scores to 100% while leaving the remaining scores unchanged. This method of adjusting is particularly appropriate when hometown scores are very nearly 100%. Casad (1974:32) has suggested that as results from intelligibility testing become so reliable that hometown scores do approach 100%, this kind of adjustment is most appropriate.

(2) Subject abilities - It could be that the subject was forgetful or unintelligent or uncomfortable in the testing situation. If this were the case we would expect these kinds of factors to effect not just the hometown test, but all tests which that subject took. The solution then would be to adjust all of a subject's (or group of subjects') scores on the basis of the score received on the hometown test. That is, the hometown score will be raised to 100% and all other scores will be raised in a comparable manner. The rationale behind such an adjustment is that no subjects should be expected to do better on an intelligibility test than they did on their own hometown test.

(3) Test deficiencies - It could be that the text on which the test was based was difficult in subject matter, that the recording was of a poor quality, that questions were improperly phrased, or that the text was segmented in inappropriate spots. If this were the case, the deficiencies in the test would affect not only the hometown scores for that test, but also all the scores for that test. The solution then would be to adjust all the scores obtained on a particular test on the basis of the score obtained by the hometown dialect. The rationale here is that no subjects should be expected to do better on a test than the hometown people did.

In the adjustments for subject abilities and test deficiencies, where not only the hometown score is adjusted but also all the other scores, there are two strategies which can be used to make the adjustment: proportional or constant. In the proportional adjustment, the adjusted score is obtained by dividing the raw score by the hometown score and multiplying by 100 to bring the results back to a percentage range. The effect is that all scores are raised by an amount proportional to the size of the raw score. In a constant adjustment, the adjusted score is obtained by adding to the raw score the difference between 100% and the hometown score. The effect here is that all scores are adjusted by adding a constant amount. As a result, the constant adjustment always yields a score greater than the proportional adjustment for scores less than 100%.

There are thus five possible methods for adjusting a raw intelligibility score:

(1) hometown,
    adjusted = 100%, if raw score is a hometown score;
             = raw score, otherwise

(2) proportional for subject,
    adjusted = (raw / hometown score for subject) x 100

(3) constant for subject,
    adjusted = raw + 100 - hometown for subject

(4) proportional for test,
    adjusted = (raw / hometown score for test) x 100

(5) constant for test,
    adjusted = raw + 100 - hometown score for test

Actually, there is no reason to believe that for any given set of data only one type of adjustment is needed. That is, it is probably closer to reality that the effect of learning curve, subject abilities, and test deficiencies could be simultaneously affecting all the results. To find the combination of adjustments which gives optimal results, however, would take the analysis beyond the techniques of correlation and regression and into the field of dynamic programming. Thus far this has not been attempted; only the effects of one adjustment at a time have been studied.

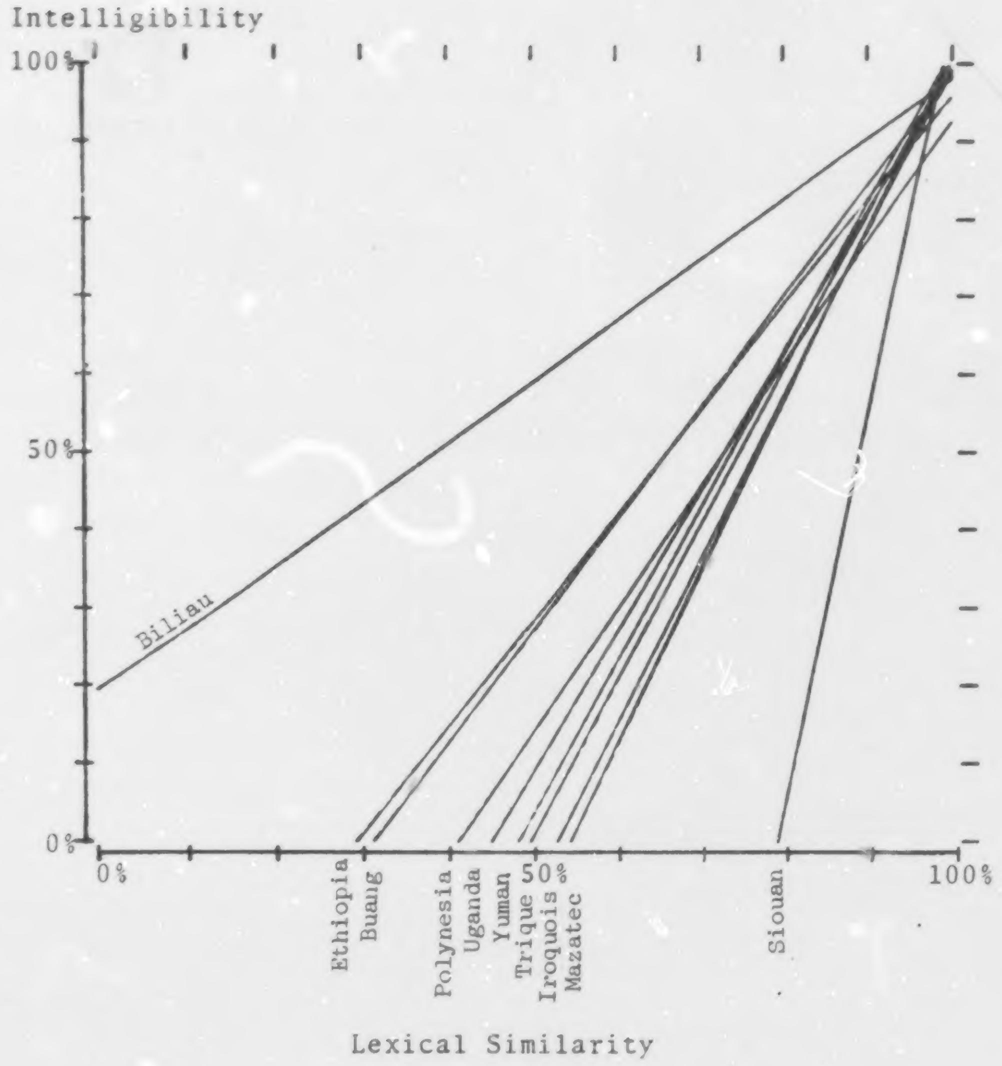
No previous investigators have come up with suggestions about which adjustments are most appropriate for what situations. Thus all five adjustments were made on all cases in the data sample in order to find the adjustment which was most appropriate for each set of data. The rationale used for selecting one adjustment as the best is explained in the next paragraph. In the listing of the raw data in Appendix 1.2, the hometown score for the subject and the hometown score for the test are listed for each data case. These values, along with the raw intelligibility score, plug into the above formulas to compute the adjusted scores. The complete set of adjusted scores is not listed in the appendix. Only one adjusted score is listed for each case. This is the score which was selected as most appropriate for the given set of data. In the description of the data sets in Appendix 1.1, the adjustment used for each set is listed.

The rationale for selecting one method of adjustment as most appropriate for a given set of data is based on two main assumptions. The first is that there is a regular relationship between intelligibility and lexical similarity. The second is that the effects of learning curve, subject abilities, and test deficiencies introduce measurement errors which perturb, not enhance, the regularity of the relationship. From these assumptions it follows that an adjustment which brings out a greater regularity is likely to be nearer the actual underlying relationship than one which reduces the regularity. To evaluate the effects of the different adjustment methods, each of the five possible adjustments was performed on each of the ten data sets. For each data set the methods were compared to find the one which brought out the most regularity from the raw data. Three criteria were used to judge this: maximizing the percentage of explained variation, minimizing the deviation from 100% of the predicted value of intelligibility for 100% lexical similarity, and minimizing the deviatioin from the mean of the value of lexical similarity for 0% intelligibility. The first has to do with regularity within the particular set of data; the second two have to do with regularity between sets of data and with a theoretical norm. Never were the three criteria met in the same adjustment method. It was therefore necessary to make a rather subjective judgement as to which adjustment gave the best combined effect. The complete set of figures on which these judgments were based and a fuller explanation of their meaning are given in Appendix 1.4 so that the interested reader can better understand and evaluate the selection process used.

In Appendix 1.5 new scattergrams for each of the data sets are plotted. This time lexical similarity is plotted against adjusted intelligibility scores. Again, the cases demonstrating an intelligibility boost from nonsymmetric social factors are plotted as "x" and the others are plotted as circles. As before in Appendix 1.3, the two regression lines are drawn in and the key statistics are listed below the scattergram.

In Figure 5.6 the regression lines for the ten sets of adjusted data with "x" points excluded are superimposed in one graph. In comparing this graph with Figures 5.1 and 5.3, two things are to be noted. First, there is a much sharper convergence of the predicting lines toward the (100%, 100%) point. Second, the fanning out of the lines at the bottom of the graph has been narrowed. The result is that the eight lines which lie in the middle very nearly represent the same underlying relationship between lexical similarity and intelligibility.

Figure 5.6 Plots for adjusted intelligibility with exclusions



intelligibility.

The details of the ten prediction lines are summarized in Figure 5.7. The format of this table is identical to that of Figures 5.2 and 5.4.

The effect of adjusting intelligibility scores can be seen by comparing Figure 5.7 with Figure 5.4, the summary for the previous stage in the analysis. The increase in percentage of explained variation is only 3.5%. Changes in the correlation coefficient, significance, and standard error of estimate are likewise minor. The significant changes are in the final two values, "Lex-100" and "Int-0". The average predicted intelligibility for 100% lexical similarity rises from 88% to 99%; the standard deviation for this value improves sharply from 11% to 3%. With adjusted intelligibility scores, the predictions therefore give a natural result -- that completely similar dialects share complete intelligibility. The variation between the prediction lines at the lower end is also reduced; the standard deviation for the point at which the lines cross the similarity axis is reduced from 40 to 25. If the two sets of data on the periphery (Biliau and Siouan) are not considered, the degree of agreement between the other eight studies stands out. The standard deviation for the crossing point is only 8.8, with the mean at 43.8% lexical similarity.

In comparing the effects of controlling for nonsymmetric social factors and controlling for intelligibility measurement error, the following can be observed. The control for social factors improves the prediction accuracy within the various studies; the adjustment of raw intelligibility scores improves the agreement of predictions between studies. In other words, the one decreases variation within studies while the other decreases variation between studies.

5.2.5 Conclusions

The goal of this analysis has been to see how well lexical similarity predicts intelligibility. The purpose has been twofold: first, to determine the relationship between intelligibility and degree of linguistic similarity, and second, to determine how well lexical similarity can function as an approximation to linguistic similarity. The main statistic which has been used to evaluate the results is the percentage of explained variation. At each step in the analysis the goal has been to explain more variation in intelligibility than was explained in the previous step by incorporating a new factor into the model to account for some of the previously unexplained variation. The final step has produced the following model to explain variation in intelligibility:

total variation in intelligibility =
    variation explained by lexical similarity +
    variation explained by nonsymmetric social factors +
    variation explained by intelligibility measurement error +
    unexplained variation

where unexplained variation includes variation due to nonlexical aspects of linguistic similarity, symmetric social relations, intelligibility measurement error not accounted for by hometown score adjustment (mainly sampling errors), and lexical similarity measurement error.

In Section 5.2.2 we found that on the average lexical similarity alone explains

Figure 5.7 Statistics for adjusted intelligibility with exclusions

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
Biliau	6	77.2	.87855	.0212	4.5	98.1	-24.6
Buang	15	65.5	.80949	.0003	15.6	98.6	31.1
Ethiopia	23	78.9	.88827	.0001	16.3	96.0	29.2
Iroquois	12	88.6	.94131	.0001	15.9	99.4	52.6
Mazatec	17	77.6	.88111	.0001	11.7	100.9	54.2
Polynesia	67	84.3	.91840	.0001	11.0	92.9	41.0
Siouan	20	79.9	.89298	.0001	14.9	105.8	78.8
Trique	11	90.0	.98174	.0001	6.5	100.6	49.4
Uganda	9	96.4	.98174	.0001	7.4	98.8	44.9
Yuman	21	99.4	.99715	.0001	2.8	101.6	48.0
Minimum	6	65.5	.80949	.0001	2.8	92.9	-24.6
Maximum	67	99.4	.99715	.0212	16.3	105.8	78.8
Mean	20	83.8	.91376	.0022	10.7	99.3	40.5
Deviation	16	9.6	.05276	.0063	4.8	3.3	25.3

65% of the variation in raw intelligibility scores. In Section 5.2.3 we found that by excluding cases in which it was suspected that nonsymmetric social factors boosted intelligibility, the percentage of explained variation was increased to 81%. We can therefore infer that the difference between these two percentages, or 16%, is the amount of unexplained variation in the original formulation which was due to nonsymmetric social factors. In Section 5.2.4 we found that if the cases which explained 81% of the variation in intelligibility were adjusted to control for some aspects of intelligibility measurement error, then the percentage of explained variation was raised to 84%. We can therefore infer that the amount of unexplained variation in the original formulation which was due to intelligibility measurement error was 3%. The decomposition of total variation is as follows:

variation due to lexical similarity	65%
variation due to nonsymmetric social factors	15%
variation due to intelligibility measurement error	3%
unexplained variation	16%
total variation in intelligibility	100%

This method of decomposing variation is called a hierarchical one, in that the components in the total variation are peeled off layer by layer. If the order in which the components are extracted is changed, the magnitude of the percentage of explained variation for each component may change slightly. For example, when the effect of measurement error is controlled for first, and nonsymmetric social factors second, the decomposition is as follows:

variation due to lexical similarity	65%
variation due to intelligibility measurement error	5%
variation due to nonsymmetric social factors	14%
unexplained variation	16%
total variation in intelligibility	100%

For the sake of interpreting the results, this latter ordering of the decomposition is perhaps more natural than the former. The former was followed in the analysis because the social factors explained a much greater proportion of the variation than did the intelligibility measurement error. By controlling for the social factors first it was possible in the analysis to select the methods of intelligibility score adjustment so as to give the most refined analysis for the final result.

In this latter decomposition, 70% of the total variation in intelligibility is explained by the first two factors, lexical similarity and intelligibility measurement error. This explanation of 70% of the variation in intelligibility has been made with recourse to only two variables, measured intelligibility and measured lexical similarity. The control for intelligibility measurement error comes only through a systematic transformation of the original measurements based on measurement of hometown scores. Thus no additional variables are measured or included in the model. The fact that by knowing only one thing about the relationship between speech communities, the degree of lexical similarity between them, we can explain the intelligibility relations between them with 70% accuracy is a dramatic result.

Many investigators have avoided the use of lexical comparison as a means of estimating intelligibility on the grounds that there are so many other factors involved: phonological similarity, grammatical similarity, semantic similarity, social relationships, political relationships, economic relationships, and geographic relationships. Nevertheless, for these ten field studies, the single factor of lexical similarity explains 70% of the variation in intelligibility in the average case. The many other factors serve only to account for the remaining 30% of unexplained variation. This does not necessarily mean that these other factors are irrelevant or of only minor importance; rather, it probably indicates that lexical similarity parallels other aspects of linguistic similarity and even some aspects of contact such as measures of social and geographical proximity.

The implication for field research is clear: lexicostatistic comparisons are a valuable tool in sociolinguistic research on communication between speech communities. This is not only because they are quick and easy, but also because they serve as reasonable estimaters of intelligibility.

The fact that the regression lines in Figure 5.6 agree to such a great extent also has important implications. Eight out of ten of the field studies point to nearly the same underlying relationship between lexical similarity and intelligibility. This suggests that it is not vain to search for a universal relationship between linguistic similarity and inherent intelligibility (that is, intelligibility based entirely on linguistic similarity and not at all on learning due to contact).

Of the two studies which do not fit the general pattern, one predicts higher intelligibility and the other predicts lower. In the Biliau study, the one which predicts higher intelligibility, the cause is definitely symmetric social relations. The two most divergent dialects in that study are only three hours' walking distance apart and there is a lot of contact between them in both directions. In the Siouan study, the one which predicts lower intelligibility, the available data do not provide an answer. The cause may lie in some aspects of linguistic similarity other than cognate percentages. If this were so, then in only one out of ten field studies did lexical similarity fail to parallel other aspects of linguistic similarity.

Raymond Gordon (personal communication), one of the investigators in the Siouan survey, suggests that the low intelligibility scores may reflect an unwillingness on the part of the subjects to give a response when they were at all uncertain. This is an interesting hypothesis which deserves further attention in future intelligibility surveys. It suggests that this is one case where socio-cultural factors in the test situation would hardly affect the hometown test (since there would be little or no uncertainty) but would affect the other tests. Therefore this kind of measurement error would go undetected by the raw score adjustment methods discussed in Section 5.2.4.

A final observation is that the results show a striking uniformity in spite of the fact that the ten studies were conducted by ten different investigators, all of whom used different methods for measuring intelligibility and different word lists and variations in technique for scoring lexical similarity. The implication here for intelligibility testing methods is that no one method is inherently better than another. Some investigators used a translation approach, some used an open-ended question approach, and others used a multiple choice question approach; some used an oral approach and others used a written approach; some used a vernacular approach and others used a common language approach. In spite of these differences, the results from study to study are surprisingly similar. This would suggest that the decision as to which kind of method to use is not based on the inherent merits of the method, but is based on the abilities of the subjects and the goals of the investigator (Section 2.3.)

The implications of this uniformity in results for lexicostatistics is that its future in synchronic research on communication potential between dialects is promising. A number of authors (for instance, McElhanon 1971:141, Hymes 1960:32) have expressed concern that lexicostatistics must undergo some precise development and standardization if results of the method are going to be valid and comparable. Their remarks are relevant mainly to the diachronic, or historical, application of lexicostatistics to questions of linguistic history and taxonomy (Simons 1977d:14-17). Here, on the other hand, we have seen that whether investigators use 100-, 165-, or 200-word lists, and whether they elicit basic or cultural vocabulary, despite idiosyncratic differences in eliciting and scoring methods, the underlying results of all methods are strikingly similar.

5.2.6 General models for predicting intelligibility from lexical similarity

In Figure 5.6 it was shown that eight of the ten field studies very nearly suggest the same underlying relationship between intelligibility and lexical similarity. The data from these eight studies are now pooled together to form one large data set. The object of this section is to investigate the possible universal relationship between intelligibility and lexical similarity as evidenced by these combined data. First two linear models are given, then seven different nonlinear models are explored. The nonlinear models offer slight improvements in prediction accuracy, but in no case is this improvement statistically significant. The final conclusion is that the data points are too scattered to permit much discrimination between different models.

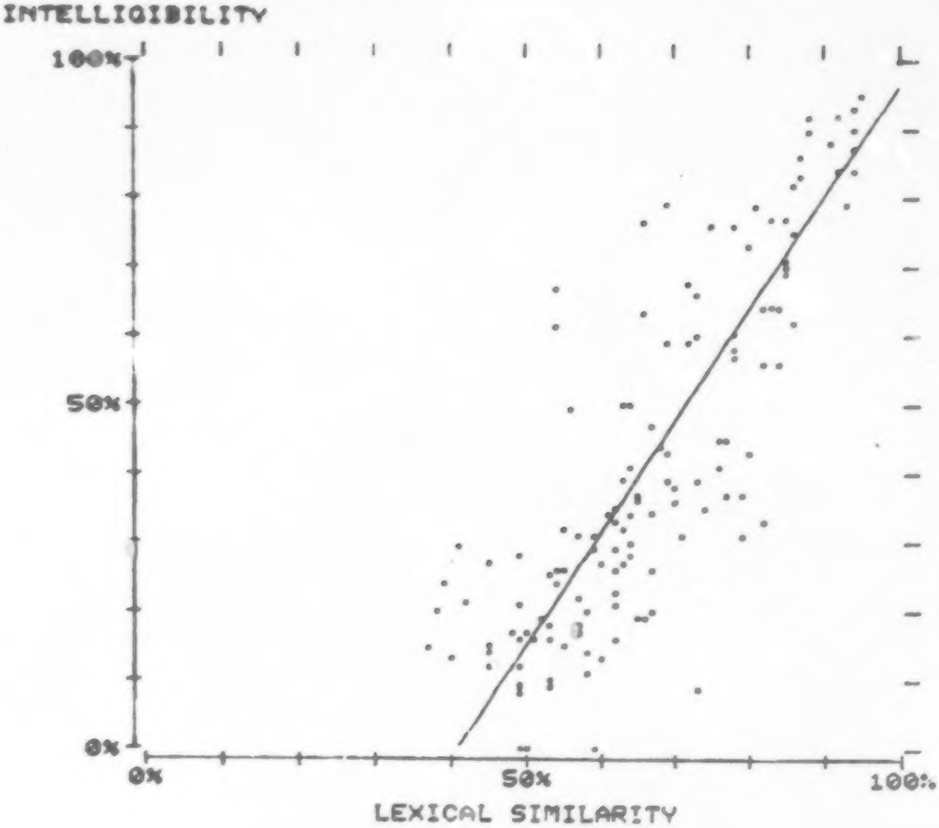
The complete pooled data set is shown in the scattergram in Figure 5.8. It contains 175 cases. Adjusted intelligibility scores are used and the points are excluded in which an intelligibility boost from nonsymmetric social factors is suspected. The straight line which best describes the relationship between the two variables is drawn into the graph. The equation for the line is written below the scattergram. Note that the model explains nearly 85% of the variation in intelligibility, and that the standard error of estimate for predictions based on the model is 13%.

The slope constant for the linear model is nearly 1.667, or five-thirds, and the intercept constant is nearly -66.67. In Figure 5.9 the linear model is simplified by rounding the constants to these values which are easy to work with and easy to remember. If the 1.667 is factored out of the -66.67, the resulting formulation makes the model more transparent as to its meaning:

$$\text{Intelligibility} = 5/3\ (\text{Similarity} - 40)$$

This model says that when the lexical similarity is below 40%, there will be no

Figure 5.8 Linear model

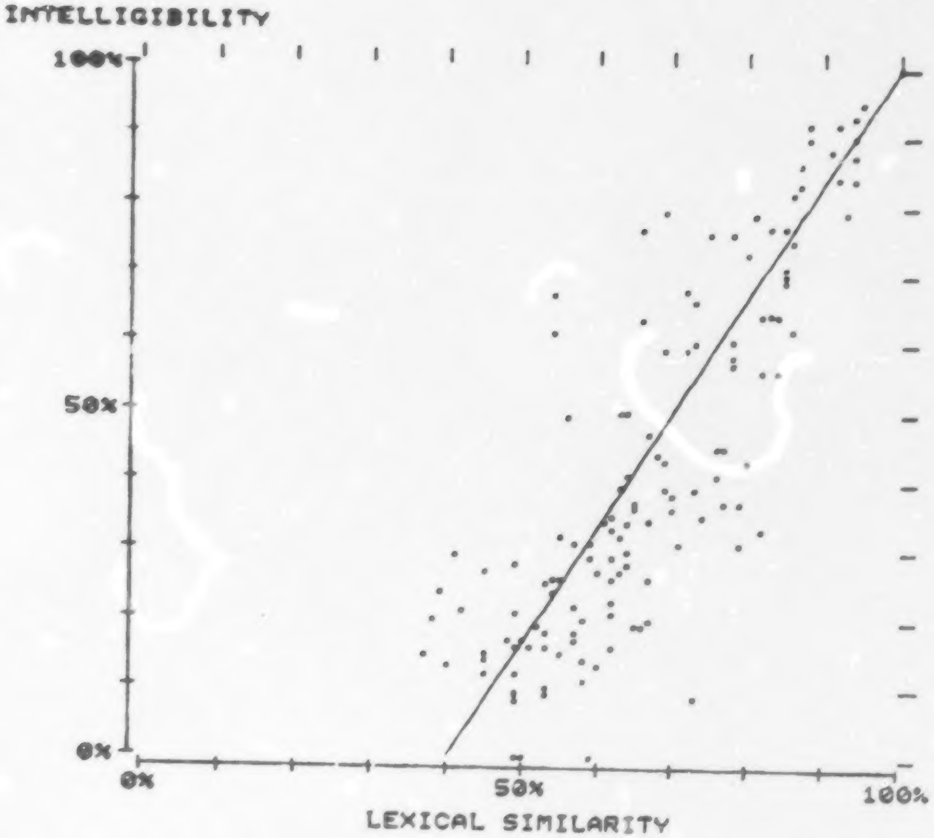


Int = 1.625 Lex - 66.12

%EV = 84.8

SEE = 13.0

Figure 5.9 Simplified linear model



$$Int = 1.667\ Lex - 66.67,$$

$$\text{or } Int = 1.667\ (Lex - 40)$$

%EV = 84.2

SEE = 13.3

understanding (it actually predicts a negative value). When lexical similarity exceeds 40%, the percentage of intelligibility is five-thirds times the amount by which the similarity exceeds 40%. Stated in another way, for every percentage point which lexical similarity increases beyond 40%, the degree of intelligibility increases by one and two-thirds per cent. Simplifying the model in this way reduces the percentage of explained variation from that of the exact model given in Figure 5.8 by less than one per cent.

There is so much scatter in the scattergrams that it is difficult to see what kind of trend the data actually suggest. One way to remove the scatter but still preserve the actual trend in the data is to plot the mean value of intelligibility for specific ranges of lexical similarity. This is done in Figure 5.10. The similarity scale is divided into segments spanning five percentage points. For instance, all the points with similarity greater than 95% and equal to or less than 100% are treated as one subset. The average lexical similarity and intelligibility for these points is computed and the point where those two values intersect is plotted. The same is done for the range of 90% to 95%, 85% to 90%, and so on until all the data points are accounted for. The plotted points of mean similarity versus mean intelligibility are connected with solid lines in a dot-to-dot fashion to give a graph of the trend in the data. This plot will appear in dashed lines in all of the graphs for nonlinear models which follow.

Now seven different nonlinear models are explored. All offer slight improvements in the percentage of explained variation over the simplified linear model. However, the greatest improvement is only four per cent and none of the models can be said to be significantly better than any of the others. This is because of the amount of scatter in the data. I suspect that as methods of measuring intelligibility, linguistic similarity, and social contact relationships are refined, the amount of scatter in future plots of this kind will be reduced and the differences between the degree of fit of the different models will become significant. The following discussion and graphs of nonlinear models are included not so much for what they reveal about the current data but as a guide for future research. The nonlinear functions were fitted to the data by least-squares techniques with a computer program. For input I specified the data and the form of the equation; the program computed the values of the constants in the equations.

A nonlinear model which immediately comes to mind is one based directly on the trend line. For each five percentage point segment on the similarity scale we could predict that the degree of intelligibility will be the mean intelligibility for that segment. This model is plotted in Figure 5.11; it is called a step function. This model gives the highest percentage of explained variation of all the models we consider (88.6%) but that is little wonder since the predictions are based directly on what intelligibility was observed to be rather than on what we might expect it to be on the basis of some general mathematical function. The model is actually quite clumsy in that the mathematical formulation of it is so lengthy and it does not seem very natural. We would expect that intelligibility would consistently increase as lexical similarity increases, not fluctuate up and down as the model in Figure 5.11 does at the lower end of the graph.

One way to approximate a nonlinear function is to use different linear functions to describe different portions of the curve. In Figure 5.12 the intelligibility curve is approximated by two straight lines. Inspection of the trend line shows that above 60% similarity, intelligibility steadily rises. Below

97

Figure 5.10 Trend line

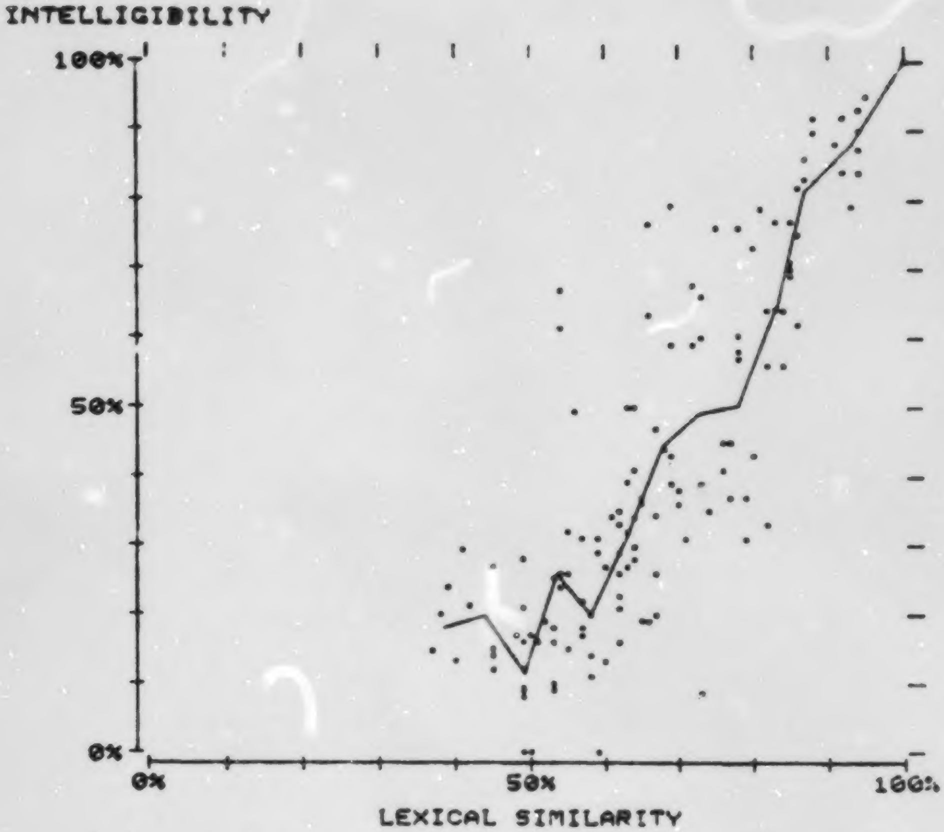
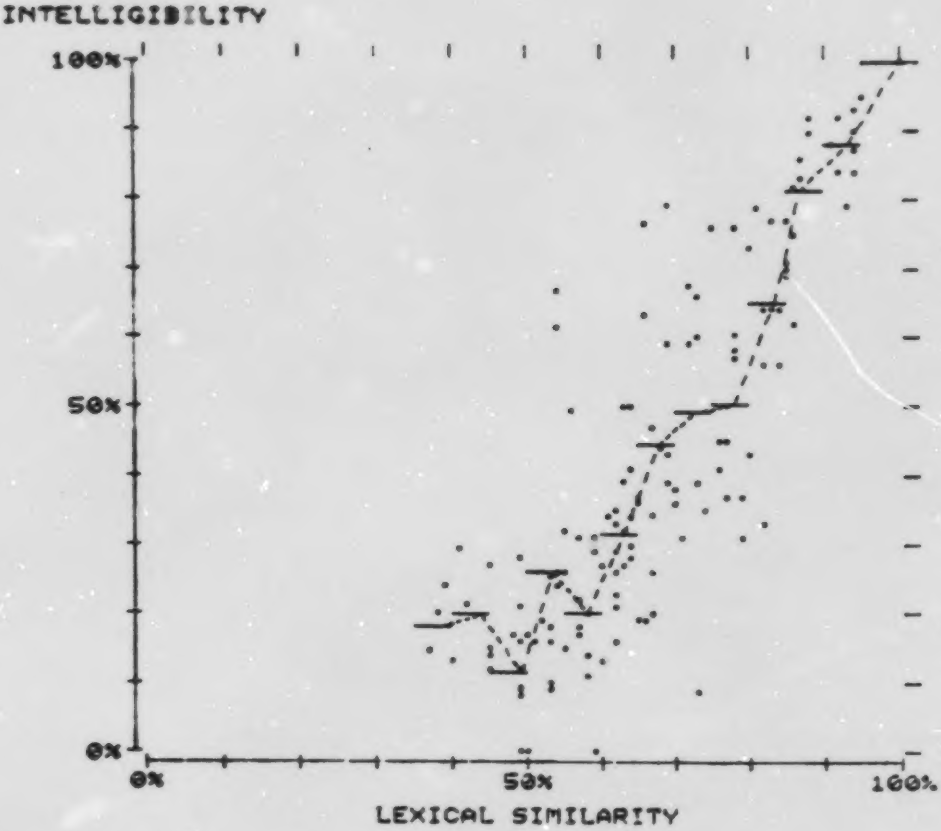


Figure 5.11 Step function

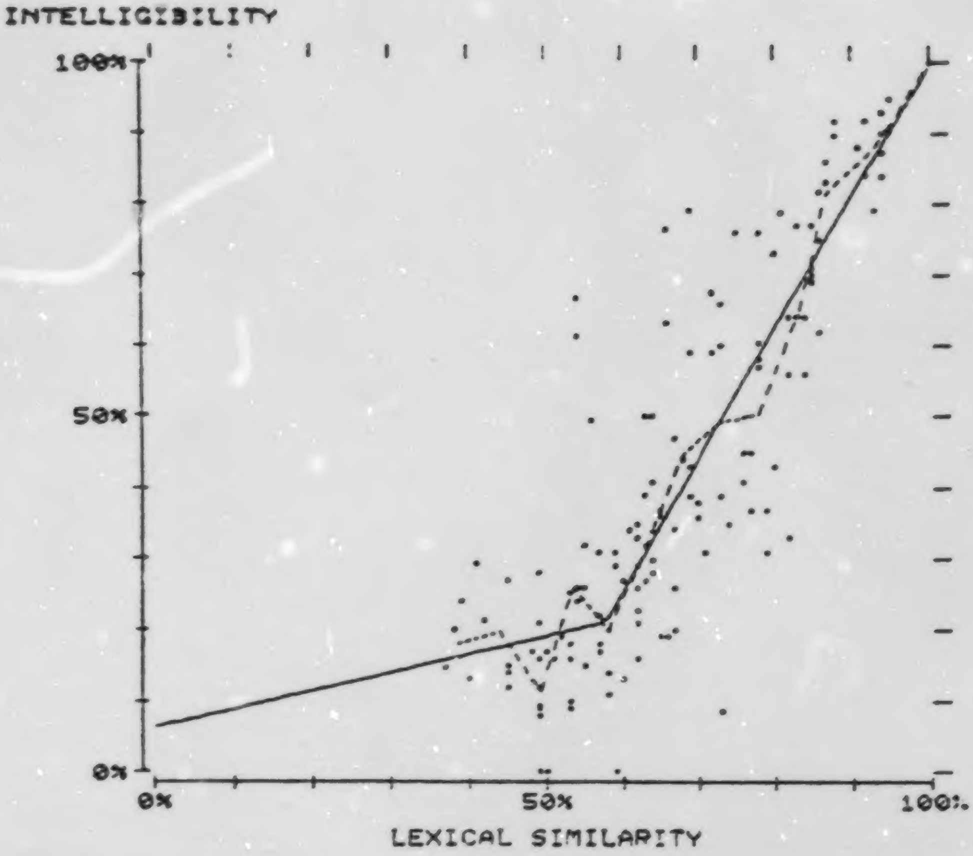


Int = 18.0, if 35 < Lex ≤ 40
= 19.8, if 40 < Lex ≤ 45
= 11.4, if 45 < Lex ≤ 50
= 26.0, if 50 < Lex ≤ 55
= 20.0, if 55 < Lex ≤ 60
= 31.5, if 60 < Lex ≤ 65
= 44.6, if 65 < Lex ≤ 70
= 49.1, if 70 < Lex ≤ 75
= 50.3, if 75 < Lex ≤ 80
= 65.0, if 80 < Lex ≤ 85
= 81.3, if 85 < Lex ≤ 90
= 88.0, if 90 < Lex ≤ 95
= 100, if 95 < Lex ≤ 100

%EV = 88.6

SEE = 11.3

Figure 5.12 Polygonal model



Int = Max(.256 Lex + 6.38, 1.85 Lex - 85.6)

%EV = 87.3

SEE = 11.9

60% the intelligibility fluctuates and shows no steady trend. A linear regression analysis was performed on these two halves of the data to find the lines which best fit each. The resulting function is called a polygonal function and the value of it is the maximum of the values predicted from the two linear functions. Note that the polygonal model never predicts the absence of intelligibility (although it could be constrained to do so without any significant loss in accuracy); even when there is no similarity, this models predicts 6% understanding. The simplified linear model in Figure 5.8 is also polygonal when we interpret it as predicting 0% intelligibility when similarity is 40% or less.

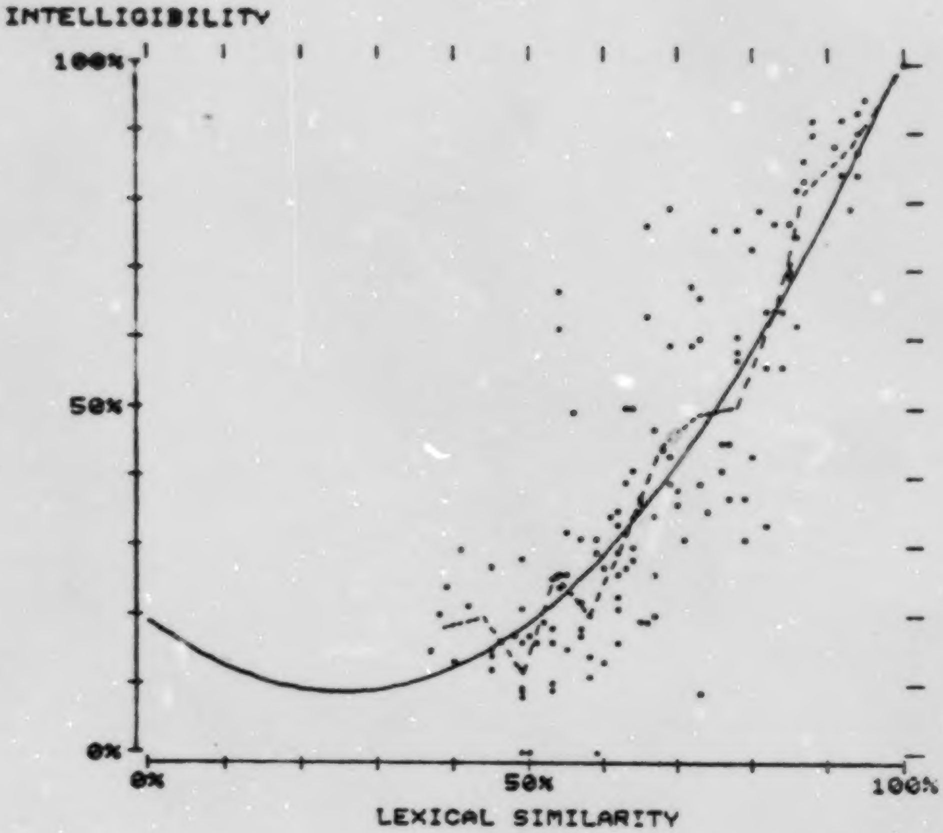
In Figure 5.13, a parabolic model is fitted to the data. Note that it too never predicts the absence of intelligibility. To correct this flaw in the model, the parabola can be constrained to reach the zero level of intelligibility. This is done in Figure 5.14 by constraining the formula for the parabola such that the intelligibility coordinate of the minimum point of the curve is zero.

In Figures 5.15 and 5.16 two exponential functions are plotted. The first is the basic exponential function. Note that it never reaches zero. Figure 5.16 plots a modified exponential function which incorporates an additive constant to bring the curve below the zero intelligibility axis.

The final model is called a logisitic, or S-curve, model. It is plotted in Figure 5.17. The logistic model is unique among all of the models considered here in that it places an upper limit on the value of intelligibility and predicts that as the upper limit is neared, increases in similarity have less and less effect on intelligibility. Whereas in all the other models, intelligibility increases at a constant or a growing rate as similarity increases, in the logistic model the rate of change slows down and levels off as intelligibility reaches its limit of 100%. This is in line with a theoretical expectation, namely the role of redundancy in dialect intelligibility. Because of the redundancy in language, listeners are able to fill in some of the items they hear that are not familiar to them. I would expect that in the range of 70% to 90% similarity, the redundancy strategy is used with the greatest benefit. In this range, an increase in similarity could be expected to give a substantial increase in intelligibility, not only because that much more is similar but because that much more can be used as a base from which to fill in that which is not familiar. Above 90% similarity, most everything would be understood so that increasing the similarity would only slightly increase the intelligibility.

Of all the models explored, the logistic seems the most theoretically satisfying. However, the current data do not give strong evidence that the relationship between intelligibility and lexical similarity is a logistic one. One problem is the degree of scatter in the data which has already been mentioned. Another factor is that the formula used represents a symmetric curve although the relationship may not actually be. The curve is symmetric around the flexion point (72%). This explains why 100% similarity does not predict 100% intelligibility in Figure 5.17. The many data points around 50% similarity pull up the curve at the low end, which has the effect of pulling it down at the high end. This shortcoming might be overcome by proposing a model which was not symmetric. This could be theoretically justified by demonstrating that different understanding processes are at work at different ends of the intelligibility scale.

Figure 5.13 Parabolic model

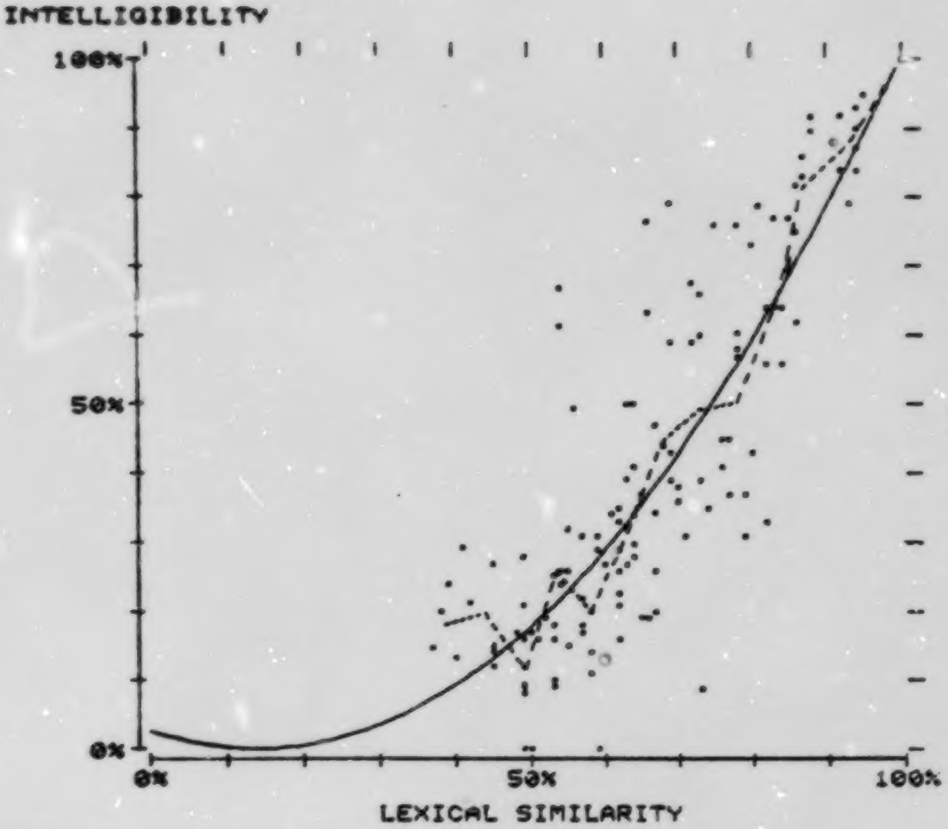


$$Int = .0165\ Lex^2 - .828\ Lex + 19.06$$

%EV = 87.1

SEE = 12.0

Figure 5.14 Constrained parabolic model

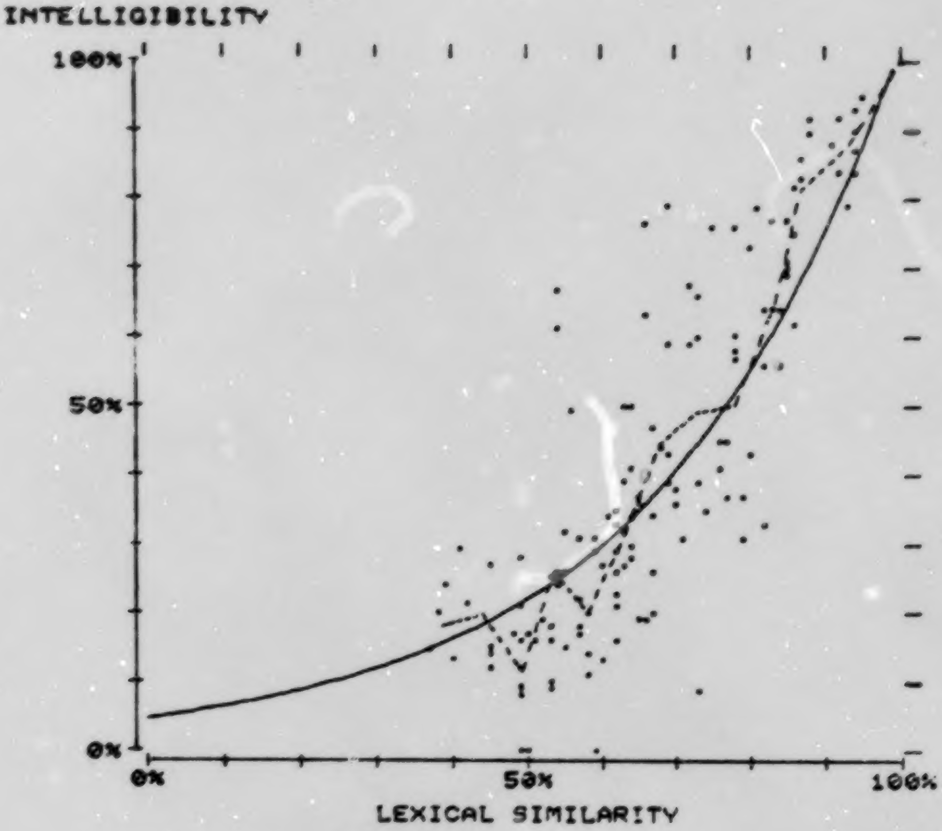


$$Int = .0135(Lex - 13.68)^2$$

%EV = 87.0

SEE = 12.0

Figure 5.15 Exponential model

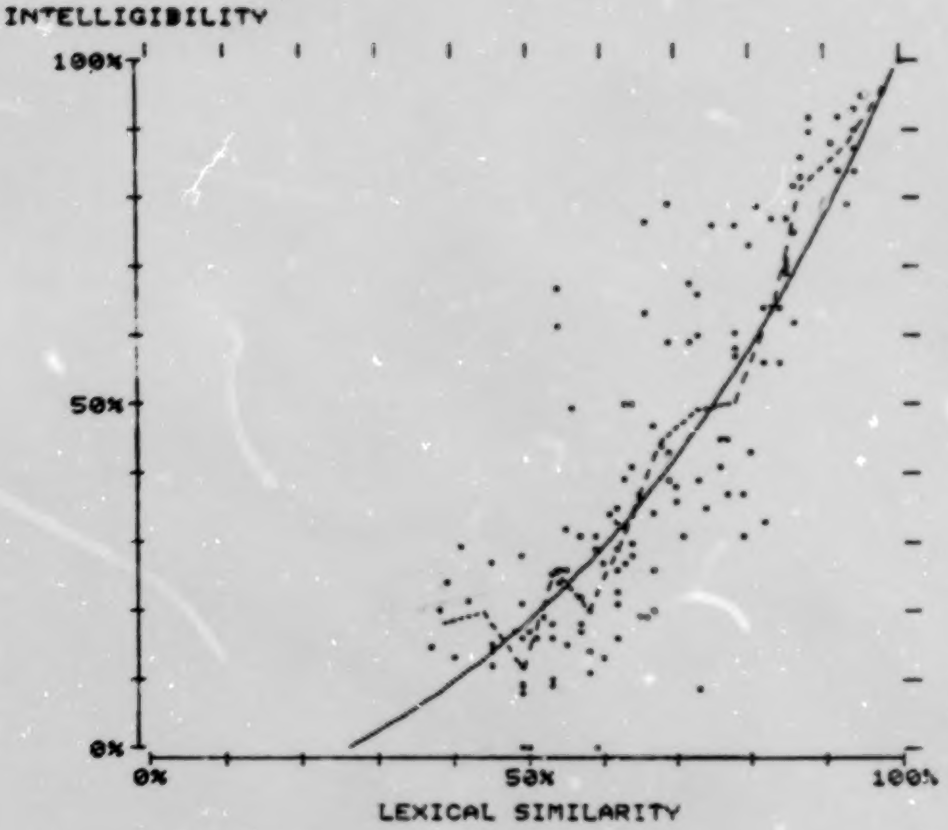


Int = $4.75e^{.0307Lex}$, where e is the natural number (2.718...)

%EV = 86.4

SEE = 12.3

Figure 5.16 Modified exponential model

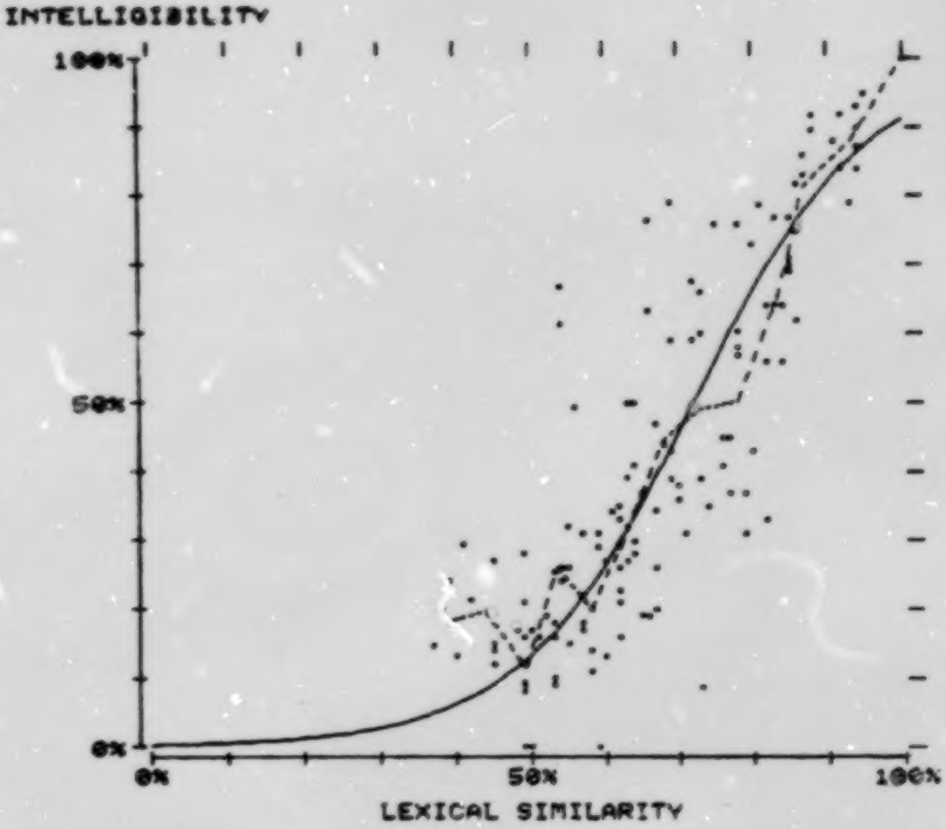


$\text{Int} = 18.75e^{.0195\text{Lex}} - 31.09$, where *e* is the natural number (2.718...)

%EV = 86.9

SEE = 12.1

Figure 5.17 Logistic model



$$\text{Int} = \frac{100}{1 + e^{-.0844(\text{Lex}-72.07)}}$$

%EV = 84.9

SEE = 13.0

CHAPTER 6

EXPLAINING COMMUNICATION: SOCIAL FACTORS

In Section 5.2.5 we saw that lexical similarity alone explains 70% of the variation in adjusted intelligibility. When presumed nonsymmetric social factors were added to the model, 84% of the variation is explained. That is, nonsymmetric social factors explain nearly half of the previously unexplained variation. If social relations had actually been measured, then symmetric social relations would have been included in the model, and even more of the variation in intelligibility might have been explained. In this chapter, this is done. The results of an empirical study of communication on Santa Cruz Island show that social factors account for most of the unexplained variation which remains after lexical similarity is controlled for.

In Section 6.1 social relations are defined and enumerated, especially with reference to their function within dialect systems. In Section 6.2 a mathematical model for explaining communication is developed; it includes both linguistic similarity and social relations. In Section 6.3 the general model is tested empirically with data from Santa Cruz Island, Solomon Islands. The results show that predictions of intelligibility made by the final model are correct 95% of the time.

6.1 The characteristics of a dialect system

A key to understanding communication patterns is to view the dialects involved as comprising a system, more specifically, a _dialect system_. A dialect system has four defining characteristics: (1) it is a _set_ of dialects, (2) the dialects are linked by relations of _interaction_, (3) the relations between pairs of dialects are defined by the _interdependence_ of all dialects, and (4) common dependence on a _center_ accounts for general patterns of communication. Each of these four characteristics is now discussed in turn.

6.1.1 A set of dialects

First of all, a dialect system consists of a set of dialects. When considering linguistic relations we concentrated on dialects as varieties of speech. Now in considering social relations our perspective turns to dialects as the groups of people who share those speech varieties.

Michael Halliday, in defining grammatical systems, is more precise about of items which comprise a system. He states that it must contain a finite number of members and that each member is exclusive of (that is, different from) all others (Halliday 1961:247). Such is true of dialect systems also where the dialect groups are the minimal members of the system.

The system must also be closed. That is, we must assume that there are no speech communities outside the system which affect patterns of communication within the system. In the real world, this is seldom true in a strict sense, except perhaps in the case of a dialect system confined to an isolated island. However, one can generally safely assume that the effects of outside speech communities are negligible when compared to the effects of inside groups. When a model is empirically tested, influence from outside the dialect system will show up as unexplained variation. If the amount of unexplained variation is negligible, the assumption of closure is justified; if it is not, the assumption may have to be reexamined.

6.1.2 Linked by interaction

A second characteristic of a dialect system is that the dialects are linked by relations of interaction. These links are social, economic, geographic, political, and ideological in nature. However, they are ultimately realized as communication between individuals in speech communities. All these different types of links are what I have thus far lumped together as "social factors". In this section different facets of interaction are explored under three main headings: channels of interaction, patterns of interaction, and measuring interaction and contact.

6.1.2.1 Channels of interaction

By channels of interaction I refer to the channels through which interaction occurs. I concentrate on the causes of interaction and classify interaction as motivated by geography, demography, community facilities, or associations.

Geography is a channel of interaction, primarily because it governs the ease of travel between speech communities. One aspect of geography is proximity. The nearer two communities are, the more likely they are to interact. This includes the likelihood of both planned interaction and chance interaction. Planned interaction occurs when a journey is made with the expressed intent of interacting with members of another speech community. Chance interaction takes place when a meeting is unplanned but occurs because members of at least one of the communities are traveling.

Other aspects of the geographic factor are terrain and routes of travel. Mountain ranges, rapid rivers, and swamps may be barriers to interaction. Conversely roads, navigable rivers, or a coast line may boost travel and interaction.

Demography, particularly the density and distribution of population, also contributes to interaction. The higher the population of a dialect, the greater the likelihood of either planned or chance interaction involving it. Not only the population, but also the density of population in the surrounding region can encourage interaction. That is, if a small speech community had a large neighbor, it would be more likely to attract interaction from more distant speech communities than if it had no neighbor at all.

Community facilities, which include for instance stores and churches, are focal points of activity where interaction takes place. These facilities attract people from other communities who come to partake of the goods and services which the facility offers. The result is interaction between members of the host community and the visitors, as well as between visitors who might come from different dialects. These community facilities are generally quite visible; they are usually located in a building or some other man made structure. Because they are so visible, they give easy but good clues to patterns of communication for the field investigator.

In most developing countries of the world it is possible to distinguish two levels of culture (Combs 1977). One is the traditional culture as practiced by the indigenous inhabitants of the land. The other is a dominant national culture which is often colonial and European in its origin. Some of the community facilities might be part of the traditional culture, for instance, a religious cult house, a traditional marketplace, or the residence of a political leader. However, in today's world, most community facilities seem to be part of the national culture. (The institutions of the traditional culture appear to be more commonly realized in the networks of associations considered next, than in specific focal points of activity.)

The community facilities of the dominant national culture may affect any aspect of life. In the case of a store or an industry, the focus of interaction is economic. In the case of a church, it is religious and social. In the case of an administrative headquarters or police station, it has to do with politics, government, or law and order. In the case of schools, the focus is on education and socialization. In the case of a hospital or clinic, it is sickness and health. In the case of a road, an airstrip, or a harbor, transportation is the focus. All of these facilities can be the site for significant contact between peoples of different dialects, and thus the location of each and the dialects served by each are important for explaining interaction.

Associations between dialects can be as important a channel of interaction as community facilities, though they are generally less visible and thus more difficult for the outside investigator to observe. Some cultural institutions realize themselves in focal locations where goods and services are obtained; others are realized in networks of associations or alliances which link dialect groups together. On the social side, marriage is one such source of interaction. When marriages occur between speakers of different dialects there are at least two relevant effects: (1) the children from that marriage usually grow up in contact with both dialects, and (2) the marriage may bind not only the two individuals, but also their whole families or lineages. The result is a channel of interaction between the groups as visits between villages are made. Adoption alliances can have similar effects in some societies.

On the economic side, traditional trading alliances can be a source of interaction. Even if the trading occurs infrequently, it can be important because it is a source of regular interaction. Perhaps the best documented example of trading alliances is the vast Kula ring off the eastern tip of Papua New Guinea (Malinowski 1922). This trading ring connected many distant islands as well as a few spots on the mainland. Although the trading occurred only once yearly, it had a profound effect on those involved and resulted in life-long partnerships between men of different islands and different languages.

Kenneth McElhanon (1970) has discussed the relation of trade routes and linguistic interaction in the Huon Peninsula of Papua New Guinea. That whole area is characterized by extremely rugged terrain. As a result, trade routes are well defined and confined to certain mountain passes. In explaining the occurrence of borrowings in lexical cognate percentages, he suggests that the borrowing occurs along the trade routes (1970:216). This is evidence for linguistic interaction along the lines of trading alliances.

Two other aspects of cultural interaction are religious and political. On the religious side, different dialect groups may interact in religious ceremonies and rituals. In Melanesia, all night dances are common (singsing in New Guinea Pidgin, dans in Solomon Islands Pidgin). These dances have their roots in the belief and ritual systems of the traditional culture; however, as islanders adopt Christianity these gatherings are beginning to take on a more purely social function. The dances are occasions for interaction between speech communities. Typically the host village invites numbers of surrounding villages (often from different dialects) to join in the dance, as well as in the giving or exchanging of vast quantities of food or valuables which may occur at the same time. On the political side, village defense alliances may span different dialects. Prominent leaders may have jurisdiction over more than their own speech community.

All of the above factors, geography, demography, community facilities, and associations, cause interaction between speech groups. In Section 6.1.2.3 some methods by which these can be measured are briefly discussed.

6.1.2.2 Patterns of interaction

In order to explain dialect intelligibility, it is necessary to distinguish between interaction and contact. To say that speakers of dialect A have frequent interaction with speakers of dialect B, suggests nothing of how well A might understand B's dialect. When they interact, they might use only A's dialect or they might use only a third language, so that A never hears B's dialect. This is where contact comes into the picture. In this hypothetical case, we would say that A has frequent interaction with B, but has no contact with B's dialect.

I define interaction to be a reciprocal, two-way phenomenon. That is, it takes place in two directions at once and in both directions it has the same intensity; A has as much interaction with B as B has with A. It makes no reference to who does the talking and it makes no reference to what varieties of speech are used. On the other hand, I restrict the meaning of contact to refer to a nonreciprocal, one-way phenomenon defined specifically in terms of the variety of speech used. By saying that A has contact with B, I mean specifically that A has contact with B's variety of speech, or dialect. The relationship is nonreciprocal and one-way because knowing how much contact A has with B's dialect tells us nothing of how much contact B has with A's dialect. It is therefore A's contact with B's dialect, not the interaction between them, which explains A's intelligibility of B.

On the basis of the contact relations involved in interaction, I propose a classification of patterns of interaction into four types: (1) balanced, (2) imbalanced, (3) rival, or (4) distant. A balanced interaction is defined as one in which the speech varieties of both participants are used to an equal extent. That is, when person A speaks, he uses his own dialect; when B speaks, he uses his own dialect; or they both could swap off using each other's dialect. For predicting communication between dialects, it is more useful to classify patterns of interaction with regard to the whole pattern of interaction between two dialects rather than in isolated conversations. Defined over a pair of dialects, balanced interaction would mean that on average both speech varieties are used to an equal extent. It could be that when speakers of A and B meet in the village of A, both speaker and hearer use the dialect of A. But a balanced interaction would also imply that if these same speakers were in village B, both speaker and hearer would use B's dialect.

Skipping to the fourth type of interaction, distant interaction, it is also straightforward and needs little comment. In such a type of interaction, the participants have had so little interaction (probably due to geographic and linguistic separation) that they are not able to use either of their own dialects. Instead, they must use a common language such as a trade language or the national language.

The second type of interaction, imbalanced interaction, is one in which contact between the two dialects is greater in one direction than in the other. This pattern of interaction is especially important to the language planner because it commonly results in nonreciprocal intelligibility and points to centers in patterns of communication. Using the example of speakers from dialects A and B, interaction would be imbalanced if both participants generally spoke A's dialect when conversing with one another, or if both generally spoke B's dialect.

The explanation of imbalanced interaction can be found in the causes of interaction discussed in the previous section. The imbalance could be due to the central geographic location of one dialect as opposed to the remote location of the other. It could be due to the large population of one dialect as opposed to the small population of the other. It could be due to the availability of goods and services at the community facilities located in one dialect and their absence in the other. It could be due to the widespread marriage, trading, or defense alliances of one dialect and the limited alliances of the other. All of these relations suggest an imbalance, with the result that in each case, the movement of people will be greater in the direction of the first dialect than in that of the second. When the movement of people is imbalanced, then we can also expect that the effects of dialect contact and learning will be imbalanced. The group which puts more effort into mobility is likely to put more effort into dialect learning. The group which is more static, is more likely to be less accomodating linguistically.

The term "prestige" has been used by other investigators to label imbalancaed relationships. However, I feel that the term is not adequate because it is not general enough: a prestige relation is only a special case of an imbalanced relation. The use of the term "prestige" dates back at least to Leonard Bloomfield. In discussing the social conditions which foster language borrowing, he suggested that there are two main factors, "the density of communication and the relative prestige of different social groups" (1933:345). Charles Hockett, in his textbook on general linguistics, devotes a section of the chapter on the conditions for borrowing to the idea of prestige. He says that the speaker must have some motive for borrowing and that two motives stand out as the most important, the prestige motive and the need filling motive (1958:404). Although these authors were speaking of prestige as a motive for language borrowing, the term has become widespread and found its way into the literature on dialect intelligibility. For instance, Ladefoged, Glick, and Criper (1972:77) state that "the percentage of words in common allows us to predict the degree of comprehension, except when questions of prestige are involved." However, the word "prestige" carries with it connotations of "esteem" and "admiration". For this reason, the term is not really appropriate in the general use it has received.

The anthropologist S. F. Nadel, in his book The Theory of Social Structure, offers a general framework in which to consider imbalanced social relations. He suggests that one of the factors which explains differential status in social systems is the relative "command over services and benefits" (1957:117). He lists some of the services and benefits an individual or group might command as: "(1) material resources and benefits; (2) social dignity (prestige, esteem, status in a hierarchical sense); (4) emotional, sensual, and aesthetic gratifications; (5) moral values (the fulfilment of duties and 'missions'); and (6) transcendental values (the 'spiritual' benefits of religion)" (1957:118). Prestige figures into this list as only one of many possible motivations for imbalanced contact. The general motivation underlying all of the above is probably need or expediency or lack of alternatives. The one group has command over something (be it material resources, prestige, special learning or skills, religious knowledge) that the other group feels a need for. When the second group goes to the first group to fill that need, imbalanced interaction is likely to occur.

The remaining pattern of interaction is one of rivalry. In this relationship, the two dialects are similar and both participants could understand and perhaps even use the speech of the other. However, because of rivalry between the two groups they avoid the use of the local dialects when interacting with one another. Instead, they prefer to use a national language or trade language which serves to deny any linguistic unity between the groups. Sometimes the rivalry emotion is one sided; one group strives for disassociation, while the other group does not. In the case of the distant pattern of interaction, the distance separating the groups is such that the participants could not use the local dialects even if they wanted to; in the case of the rivalry pattern, both participants could use the local dialect, but at least one does not want to.

Hans Wolff's paper on "Intelligibility and Inter-ethnic Attitudes" (1959) gives examples of the rivalry pattern of interaction. For instance, until recently it was generally agreed by speakers of Urhobo and Isoko dialects of southwestern Nigeria that the two dialects were mutually intelligible. However, Wolff reports that lately the Isoko speakers are claiming otherwise. He states that "this claim has coincided with Isoko demands for greater self-sufficiency" (1959:37).

Attitude is a term which often enters into discussions of explaining communication (Wolff 1959, Casad 1974:185-188, Callister 1977, Collier 1977). Therefore it would be good to clarify the position of attitudes in explaining communication. By attitudes I am referring to feelings one group might have toward another, feelings such as friendliness or animosity, esteem or scorn, trust or suspicion. I feel that attitudes are not so much a direct factor in explaining degree of intelligibility as they are in explaining patterns of interaction. That is, attitudes affect patterns of interaction, which in turn affect intelligibility. Casad's model also reflects this view (1974:184-186). Thus, if contact, which factors out the components of interactions, is measured and plugged into a model to explain communication, instead of using reciprocal interaction in the model, then attitudes have already entered into the contact factor and do not play a separate role in the model. However, when contact is also predicted (see Section 6.2.2), then attitude becomes more important. Ultimately, attitude probably has a bigger role to play in determining the acceptability of materials written in one dialect to speakers of another dialect than it does in explaining one group's comprehension of another.

6.1.2.3 Measuring interaction and contact

One way to gather information about patterns of interaction is to ask about them directly. For instance, go into a speech community and ask, "When you meet someone from that other community, do you each speak your own dialects, do you speak his dialect, does he speak yours, or do you both use the trade language?" If both use their own dialects, a balanced relationship is implied. If they used one dialect exclusively, the relationship is imbalanced in the direction of that dialect. If they use a third language, distance or rivalry is indicated.

The methods described in the rest of this section refine and validate these initial findings by uncovering the channels of interaction and by estimating the degree of contact in each direction of the interaction. Many lines of investigation are suggested. Not all will be appropriate for every situation, and there will be exceptions to every trend I suggest. The approach I have used in the field is to follow up lines of investigation which seem fruitful for the given situation, and then perform a statistical analysis on the findings (Section 6.3) to discover what aspects of interaction and contact best explain communication.

As suggested in the preceding section, the one-way contact which a dialect has with another is more important in explaining communication than the two-way, reciprocal interaction between the two dialects. Thus, the methods described here concentrate on the observation and measurement of one-way contact rather than two-way interaction. The presentation here is brief; the paper by Sandra Callister (1977) on sociolinguistic approaches to dialect surveying in Papua New Guinea is probably the best guide presently available for formulating questions and presenting the results of an investigation of contact relations. She lists many more possible questions than I do here, discusses the different ways a question can be phrased, and illustrates the method of presenting results in a matrix. Delbert Miller's _Handbook of Research Design and Social Measurement_ (1977) may also be helpful. It is an exhaustive guide to sociometric methods in general.

The measurement of geographic factors is fairly straightforward. Distances can be measured from a map in miles or kilometers. A more meaningful measure is perhaps the traveling time between speech communities expressed in hours or minutes or perhaps even days. Traveling time takes into account some of the geographic barriers to interaction such as mountain ranges, as well as some of the boosters of interaction, such as roads or navigable rivers. The raw distance measurement is a two-way, reciprocal predictor of interaction. In Section 6.1.3 the concept of measuring distance relatively with respect to the dialect system rather than against an absolute measuring scale is presented. This has the result of giving a geographic estimate for one-way contact. This technique is illustrated with the data in Appendix 2.1.3 and Section 6.3.4.

Population can be measured or estimated by means of census techniques. The population of one group relative to a second gives a one-way estimate of contact. A measure of relative population is computed by dividing the population of the second group by that of the first. A score greater than one indicates that the second group is larger and that contact is likely to be imbalanced in the direction of the second dialect. A score less than one suggests an imbalance toward the first dialect. Note that the population of the first relative to the second is different than the second relative to the first, thus relative population estimates one-way contact.

Perhaps a more meaningful measure is relative density of population. Such a measure takes into account the location of neighboring villages and the possible effect they have on attracting interaction from other groups. To measure population in terms of density, rather than actual number of inhabitants, has the effect of measuring population with respect to the system rather than in isolation (see Section 6.1.3). This technique, is illustrated with the data in Appendix 2.1.4. The results for the Santa Cruz Island data show that density of population is a slightly better predictor of intelligibility than population (Figure 6.13), but not significantly better.

In measuring the contribution of community facilities to contact, the first step is to plot the location of the facilities within the area of the dialect system on a map. Much of this information may already be on existing maps. Much of it can be gathered from the simple observations of the investigator. To ensure a thorough job, however, it is generally necessary to ask questions in the villages to determine where each of these facilities is located. Some of the facilities to take note of are: churches, schools, stores, markets, clinics, hospitals, government offices, police stations, plantations, factories, truck or bus depots, airstrips, and harbors. After these data are collected a simple measure of the relative command over services of dialects in the system can be obtained by totaling up the number of facilities in each village or dialect area. One could then hypothesize that the direction of contact will be from the dialects with fewer facilities to the ones having more.

A more refined measure can be obtained by determining the domain, or area of influence, of each community facility. This is done by asking at each village where they go to obtain the goods and services they require. That is, at a particular village ask, "Where do you attend church?", "Where do you go to market?", "Where do you go to buy store goods?", "Where do your children go to school?", and so on. A method of tabulating the data is presented at the end of this section.

Gathering data on cultural associations is more difficult because it requires some investigation of the cultural traditions in the area. Marriage is probably the easiest kind of association to study. One approach to studying marriage ties is to treat them as an indicator of two-way reciprocal interaction. In this approach one notes the presence of ties between a pair of speech communities without regard to the direction the tie might take in terms of residence of the married couple. The most simple way to question and record responses is to use yes-no questions. For instance, "Is there anyone from this community who is married to someone from that community?" The answers will be yes or no and these can be recorded in a table as ones or zeros. For the opposite extreme of complexity, one could take a census approach and count or estimate the actual number of marriages that link each pair of villages. This would involve talking to every couple in a village or to a representative sample and find out where the husband and wife are from. A level of complexity which is midway between and which is probably the best for these purposes is to record responses in some scale of degrees. In such an investigation the question asked would be, "How many marriages are there between a person from this community and a person from that community?" The investigator could judge the response and score it on a three level scale such as "no marriage ties", "some ties", or "many ties". Or a scale which approximates the number of marriages could be used, such as the following four level scale: "zero", "one or two", "three to five", or "six or more". Such scaling approaches are used when the data are not so reliable and exhaustive that the investigator can be sure of the complete accuracy of the subjects' responses.

By noting the place of residence of the married couples, the result is a more refined method of measuring marriage ties as one-way indicators of contact. Such an approach should be preceded by some investigation into the marriage customs of the people to find out if it is customary for the couple to live in the community of the wife, of the husband, or of their own choosing. An understanding of the land and inheritance rights people retain if they move away and how they keep claim to them active might also be relevant. Investigation of the kinds of contact which result from marriage ties would also be helpful, such as patterns of visiting between families or the relationship of marriage ties to trading alliances and participation in ceremonies. In treating marriage in a one-way fashion, a distinction must be made between the village of residence and the village from which the spouse comes. The kinds of questions used in investigating two-way marriage ties are refined by incorporating place of residence. If place of residence is strictly prescribed as being the original community of the husband or of the wife, then the questions can be phrased in terms of sex rather than place of residence. For instance, "How many women from here have married a man from that community?" Methods of tabulating results are discussed at the end of this section.

Measuring other kinds of associations, such as trading, ceremonial, or political alliances, requires preliminary investigation to discover the nature of the interaction. Once the presence of a certain kind of interaction has been established, it is possible to formulate questions that could be asked in the dialect survey. "Who?" questions can be used to establish the fact of interaction. "Where?" questions can be used to determine the direction of contact. "How many?" and "How often?" questions can be used to estimate the degree of contact in each direction.

The above methods are designed to gather data on interaction and contact. Data on attitudes can be gathered by changing the perspective of the questions. Rather than asking about the facts of past contact, ask about preference for contact with one group as against another or ask for a value judgement concerning another village. Callister (1977:201-2) and Collier (1977:260) both give lists of possible questions to use. Miller (1977) describes methods of developing sociometric scales and indexes which could be used to develop schemes for assessing interdialect attitudes.

For all of the methods in which contact relations between speech communities are investigated, the best way to organize and tabulate the data is to put them in a two-dimensional matrix. To simplify the comparison of results from different lines of investigation, all such matrices should be consistently labeled with respect to the ordering of dialects and the orientation of the two dimensions. The data matrices in Appendix 2.1 exemplify this kind of consistent labeling and should be referred to for examples as these principles of labeling are discussed in the next two paragraphs.

The dialects should be listed in an order which causes the values for highest contact to cluster along the diagonal of the matrix and the values for lowest contact to occur on the edges. Ascher and Ascher (1963) describes an algorithm which orders matrices in this way. With this arrangement dialects which are adjacent in the ordering have a high degree of contact and those which are separated have lower degrees. In general the optimal ordering will be close to a geographic one. Alphabetical orderings are to be avoided, because they fail to bring out the natural ordering relationships which the data values themselves imply.

All matrices should be labeled with the same orientation of the two dimensions. I have adopted the convention of labeling the dialects along the left hand side (that is, the rows) as the origin and the dialects along the top (that is, the columns) as the destination. In this way the movement implied in the contact relations is read from left to right in the matrices. The following descriptive labels for the dimensions can be used. For relative distance, the labels can be "from" and "to". For community facilities they can be "domain" (or "users") and "location". For marriages they can be "place of origin" and "place of residence", or simply "women" and "men" if the prescribed pattern is one of residence in the husband's community. For ceremonials they can be "visitors" and "host". "Origin" and "destination" are another pair of labels that can be used to describe movement of people. Since the contact of the origin group with the destination group is what predicts the origin group's intelligibility of the destination group's dialect, in an intelligibility matrix "hearer" corresponds to origin and "speaker" corresponds to destination. After labeling, each of the contact measurements is written into the matrix cell where the row for the origin group intersects the column for the destination group.

These methods in which a two-dimensional matrix is filled in yield a large number of data points. If there are *n* dialects, then there will be *n*-squared possible measurements of contact between them. However, the relationships within the whole dialect system can be summarized in terms of *n* measurements of the relative attraction and motivation of dialects by summing the rows and columns of the matrix. When the rows represent origins and the columns represent destinations, then the sum of a row gives a measure of the overall motivation of that dialect group to travel and make contact; the sum of a column gives a measure of the overall attraction of that dialect. Comparing the sums of the rows gives an indication of the relative motivation of the dialects within the system to make contacts; some will show an outgoing nature while others will show a stay-at-home nature. Comparing the sums of the columns indicates relative attraction; it will become clear which dialects attract a lot of contact and which do not. This technique is used with the Santa Cruz data in Appendices 2.1.8, 2.1.9, and 2.1.10.

6.1.3 Relations defined by interdependence

A third characteristic of dialect systems is that the relations between pairs of dialects are defined by the interdependence of all dialects. This interdependence characteristic of systems is one which has been recognized and used by linguists in defining grammatical systems. Halliday states that "if a new term is added to the system, this changes the meaning of all the others" (1961:247). Kenneth Pike defines system as a group of two or more units which enter into each others' definitions (Pike and Pike 1977:139). For dialect systems this principle is realized in at least two ways. These may be summarized in the observation that (1) the measurement of distance is relative to the system, and (2) the learning effect of contact is cumulative over the whole system.

We are used to measuring distance in absolute terms. For instance, in measuring geographic distance we use absolute, universal units such as miles or meters. In measuring linguistic distance we might use a standard measure such as percentage of lexical forms which are noncognate. These absolute measures are helpful when the observer stands outside the system and attempts to measure distance. However, when the observer stands inside the system as though he were a participant, the perception of distance begins to become relative. That is, the distance from one point to another is perceived in relation to the distance of that point to all others in the system. This distinction between the perspective of one standing outside the system and that of one looking from the inside as a participant is like Pike's distinction of the etic and emic perspectives (Pike 1967:37ff, Pike and Pike 1977:483). For instance, when growing up in California I gained a view of United States geography in which Chicago is situated midway between the East and West coasts. However, when measured against an absolute scale, Chicago is three times further from San Francisco than it is from Washington, D. C. As another example, Americans generally perceive South America as being located directly south of the United States. However, when measured against the absolute scale of global longitude, the whole coastline of Chile turns out to be east of Washington, D. C.

The data from Santa Cruz Island (see Sections 6.3.4 and 6.3.5) give clear evidence that distance measured with reference to perspective from within a dialect system provides a much better explanation of communication than distance which is measured by an absolute scale and makes no reference to the system.

Figure 6.1 gives a map of Santa Cruz Island illustrating this point. The map illustrates the measurement of distance relative to two villages, Mbanua (BAN) and Nanggu (NNG). Mbanua is the geographic center for this dialect system. It is defined as such because it is, on average, nearer to all the other dialects in the system than any other dialect. The loop which surrounds Mbanua is drawn at a radius equal to the average distance from Mbanua to all other speech communities in the dialect system. This distance is 180 minutes, or 3 hours, of traveling time. One way of interpreting this loop is that from the standpoint of Mbanua, half of the people on the island live within the loop, the other half live outside of it. Nanggu, on the other hand, is the most peripheral village in the system. The average distance from Nanggu to all other dialects is higher than that for any other community. Again, an arc is circumscribed around Nanggu at a radius equal to the average distance to all other dialects. This distance is 588 minutes, or nearly 10 hours, of traveling time.

The map in Figure 6.1 should illustrate the relative nature of distance on Santa Cruz Island. A traveling distance of 180 minutes is a completely different thing from the perspective of Nanggu than it is from the perspective of Mbanua. A villager from Mbanua can meet half of the people on the island by traveling 180 minutes from home. However, if a villager from Nanggu travels 180 minutes he has barely left home and still has a long way to go to meet anyone outside of the neighboring village of Mbimba. My hypothesis is that for using distance to predict interaction, the average distance to all other dialects in the system, that is, 180 minutes for Mbanua or 588 minutes for Nanggu, are roughly equivalent distances from the perspective of the respective villages.

he hypothesis receives support from the analysis in Section 6.3.4 where it is shown that geographic distance measured by an absolute scale explains intelligibility with 67% accuracy while geographic distance measured by a relative method explains intelligibility with 83% accuracy. Stated in another way, absolute distance explains intelligibility with 33% error while relative distance explains intelligibility with 17% error, only half as much.

The second aspect of interdependence is seen in the way that the learning effects of contact are cumulative. When trying to explain communication, the degree to which one dialect understands the speech of another does not depend simply on its contact with that dialect. It depends also on the first dialect's contact with all other dialects. This is because those dialects also bear some similarity to the original target dialect. Therefore, the effects of contact with all other dialects has a contribution to learning about the speech of another dialect.

An example of this principle can also be seen from the data of Santa Cruz Island. In Figure 6.2 another map of the island is reproduced. In this case the lexical similarity between four of the dialects is indicated -- Neo (NEO), Lwowa (LWO), Mbanua (BAN), and Nooli (NOO). The lexical similarity between Nooli and Neo is only 59%. This is well below the level of similarity for which we normally expect full understanding. Swadesh suggested that 81% similarity correlated with the lower limit of full intelligibility; many investigators in Papua New Guinea have

Figure 6.1 Relative distance on Santa Cruz Island

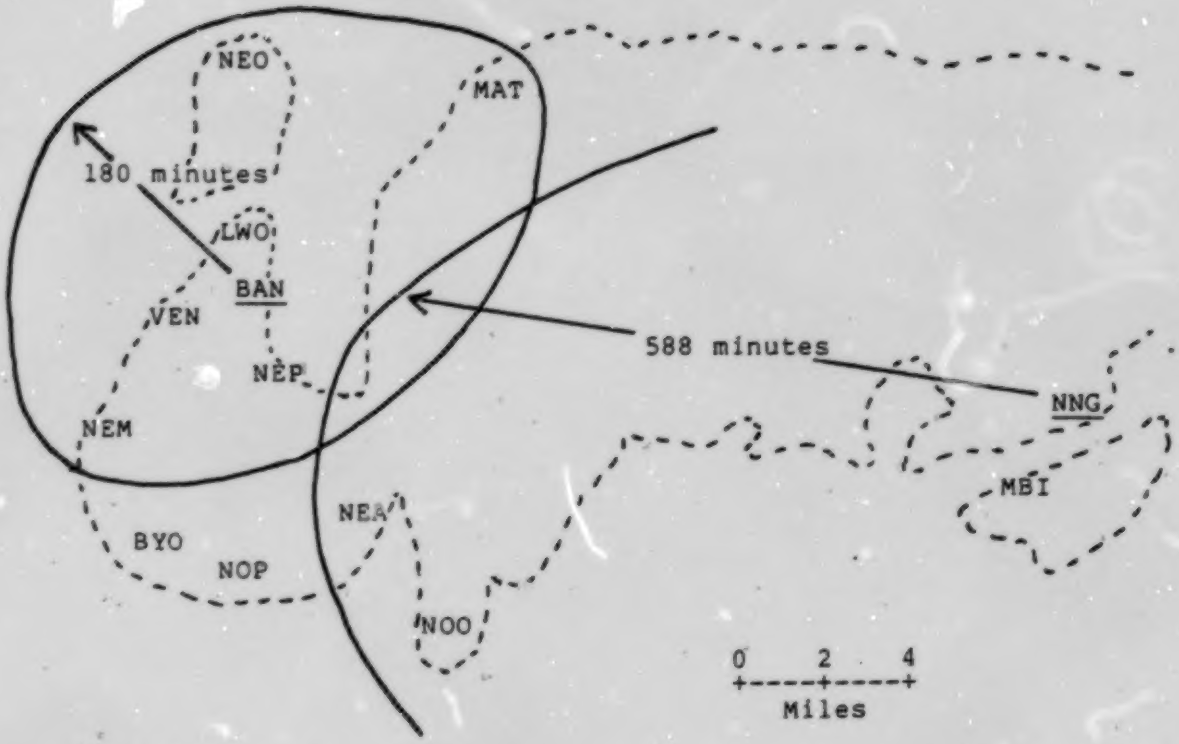
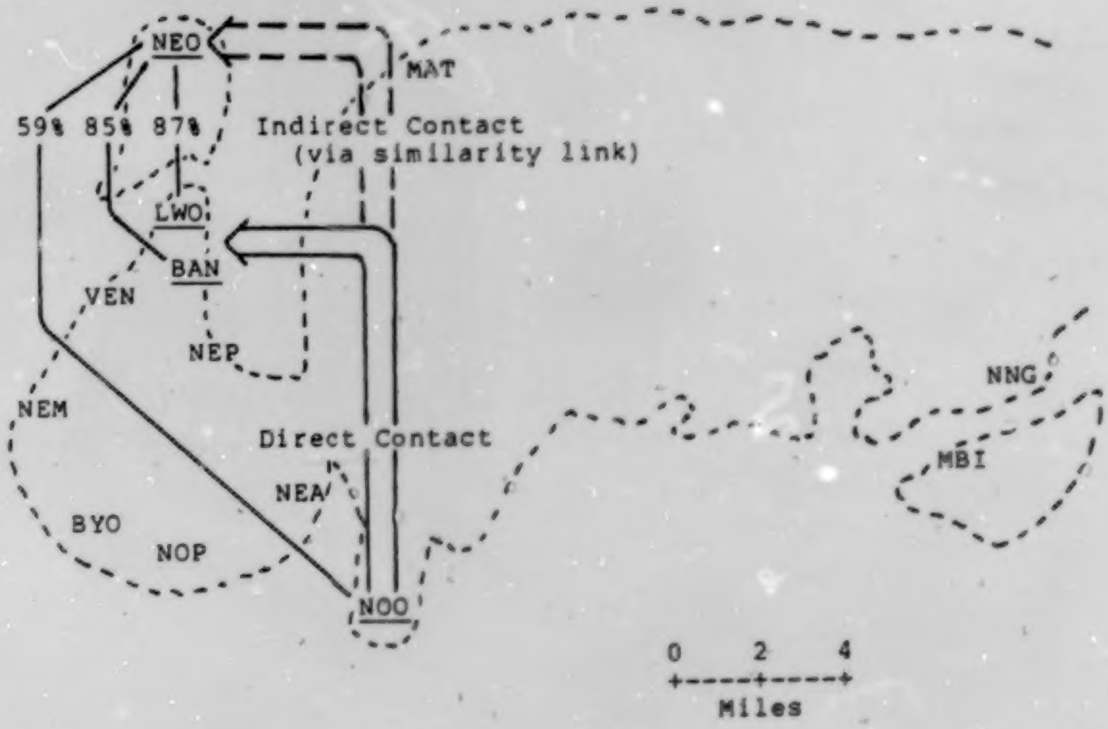

NEO
MAT
180 minutes
LWO
BAN
VEN
588 minutes
NEP
NNG
NEM
MBI
NEA
BYO
NOP
NOO
0 2 4
Miles

Figure 6.2 Indirect contact on Santa Cruz Island


NEO
MAT
59%
85%
87%
Indirect Contact
(via similarity link)
LWO
BAN
VEN
NEP
NEM
Direct Contact
NEA
BYO
NOP
NOO
NNG
MBI
0
2
4
Miles

considered 70% or 75% more realistic for that (McElhanon 1971:134-5). (Note that the simplified linear model in Figure 5.9 of Section 5.2.6 indicates an expected intelligibility of 32% when lexical similarity is 59%.) However, intelligibility tests showed that Nooli had full understanding of the speech of Neo. The parallel data which were collected on interaction relations indicate no relations of direct contact between the two villages. On the basis of this fact alone, we would then expect that Nooli's contact with Neo would not be sufficient to boost intelligibility to the full level. It is the interdependence principle which explains the presence of intelligibility. The data on interactions do indicate that Nooli has contact with Lwowa and Mbanua. This then accounts for the full intelligibility measured for Nooli on the speech of Lwowa. Note now, that Neo is 87% similar to Lwowa and 85% similar to Mbanua. This indicates that Nooli, through its contact with Lwowa and Mbanua (as well as the other dialects in the same vicinity), has learned up to 87% of the speech of Neo. Thus it is conceivable that without any contact whatsoever with Neo they would be able to understand that dialect at a level predicted by 87% lexical similarity. Actually, Neo shows some similarity to all other dialects with which Nooli has contact. Thus, through learning to understand differences from other dialects, they have at the same time learned to understand many aspects of the speech spoken at Neo. Therefore, the absence of direct contact between two speech communities is not sufficient evidence to discount learning between the dialects. The learning of another dialect is actually a function of contact with all other dialects and the similarity of those dialects with the target dialect. Gillian Sankoff observes this same phenomenon among the Buang of Papua New Guinea (1968:184, 1969:848).

6.1.4 Common dependence on a center

A fourth characteristic of dialect systems is that the relations between dialects are not random; they are subject to the common influence of a center. The solar system gives a good example of this property. The motion of the planets can be understood only in terms of the common gravitational pull of the sun. The motion of a moon within the solar system can be partly explained through the force exerted by the sun but requires the introduction of a second force, the gravitational pull of the host planet, to explain the small orbits which are superimposed on the huge orbit around the sun.

Dialect systems, too, are characterized by these common and central forces which explain the overall pattern of interaction. In Section 6.12 the imbalanced pattern of interaction between two dialects was discussed. When many such pairwise relations are viewed simultaneously, then an overall pattern of a single dialect dominating interaction with the surrounding dialects may be seen. This kind of dominance (attraction) defines centers, and thus dialect systems are viewed as centered systems. The center within a dialect system is the primary force in explaining patterns of communication within that system. As in the solar system, there may also be secondary (or even tertiary, and so on) centers. These subsidary centers would be used in addition to the primary center to explain relations in a specific subsystem of the whole.

A center is defined by recourse to a number of factors. For a given system, the communication center would be the dialect most widely understood. The linguistic center would be the dialect having the highest average linguistic similarity to all other dialects. The geographic center would be the dialect having the lowest average distance to all other dialects. The demographic center would be the dialect having the greatest population. The center with respect to community facilities would be the dialect having the greatest collection of facilities. The center for cultural associations would be the dialect attracting the greatest number of married couples, attracting the greatest number of people to ceremonials, having the greatest concentration of traditional wealth, or having the greatest political power. All of these factors contribute to defining a central dialect. An examples is given in Section 6.3.2 where the central dialect for Santa Cruz Island is determined. For three other examples of defining a central dialect and a discussion of the general topic, see the paper by Joy Sanders (1977). In Section 6.3.5 (Figure 6.13) the hypothesis that relations to the central dialect define relations over the whole system is tested. The predicting models thus derived are 88% accurate on average.

6.2 A general model for explaining communication

In this section a general model for explaining communication is developed; in Section 6.3 it is tested with data from Santa Cruz Island. Many possibilities for filling in the general model are suggested; at this point it is too soon to propose which are best. Therefore, in Section 6.3 many of the proposals are tested against the field data and the results are reported. These serve to indicate the potential of the general model.

The model is developed in two parts. First, Section 6.2.1 discusses the relationship between linguistic similarity and contact in predicting intelligibility. A model involving those three variables alone is given. Second, Section 6.2.2 concentrates on the contact variable and develops a model for predicting values of the contact factor to use in the main formula.

6.2.1 Predicting intelligibility

The basic model suggested in Section 4.3 states that intelligibility has two components, a similarity-based component and a contact-based component. Another way of saying the same thing is that intelligibility is based on both linguistic factors and social factors. In mathematical terms, one would say that intelligibility is a function of linguistic similarity and contact. That is,

$$I = f(L,C)$$

where I = intelligibility,
L = linguistic similarity, and
C = contact

The goal of this section is to specify the manner in which these two variables interact to explain intelligibility.

All previous attempts to specify a model for intelligibility have suggested that the function relating linguistic similarity and contact is an additive one (Casad 1974:191, Stoltzfus 1974:46, Collier 1977:256). That is,

$$I = f(L) + g(C)$$

This model states that intelligibility is equal to the effect of linguistic similarity, or the effect of contact, or the sum of both. When there is no contact, the C factor is zero and intelligibility is based strictly on linguistic similarity. When there is no similarity, the L factor is zero, and intelligibility is based strictly on contact. When there is both similarity and contact, intelligibility is the sum of their effects. When contact is favorable, the effect will be a boost in intelligibility above the level predicted by linguistic similarity alone. When contact is not favorable, the C factor could have a negative value which would have the effect of limiting intelligibility to a level lower than that expected on the basis of similarity. (This model therefore accounts for the cases reported by Wolff 1959.)

This model has the disadvantage that it puts no ceiling on the possible effect of contact. If the similarity and contact factors were both high, a percentage of intelligibility beyond 100% would probably be predicted. This is because the model specifies no interaction between the variables; a given degree of contact increases intelligibility by the same amount regardless of the degree of similarity. This cannot actually be the case, however. For instance, assume that 75% similarity predicts 50% intelligibility when there is no contact and that X amount of contact raises intelligibility by 40% to 90%. But suppose that 90% similarity predicts 80% intelligibility when there is no contact, then X amount of contact cannot raise intelligibility by another 40%. By definition, the degree of intelligibility cannot exceed 100%. Therefore, the degree of understanding which a certain amount of contact brings about must be restricted by the amount of improvement which is still possible.

This refinement to the model can be formulated as follows:

$$I = f(L) + g(C(100-L))$$

Here linguistic similarity is measured as a percentage. The value (100 - L) then gives the percentage of non-similarity. This model suggests that the learning (and thus intelligibility boost) brought about by the contact factor is limited to that portion of the language which is not already similar.

In order to use least-squares regression techniques to test the model, the model would be reformulated as follows:

$$I = b_0 + b_1L + b_2C(100-L)$$

Multiple regression analysis would then yield values for the three b constants in the formula.

The techniques of least-squares regression analysis are not appropriate for the data from Santa Cruz Island because step funtions rather than continuous functions are required to predict intelligibility (see Section 6.3.3). Therefore, another formulation of the basic model is tested in Section 6.3. It is as follows:

$$F = L + C(100-L)$$
$$I = f(F)$$

In the first place, linguistic similarity and contact are combined directly (with no weighting factors or additive constants) to predict the "linguistic familiarity", or F. The familiarity is a percentage estimate of what portion of the dialect of the speaker is familiar to the hearer, either through similarity or contact or both. In order to prevent F from exceeding 100%, the C factor must be scaled to a range of zero to one. As long as L ranges from 0% to 100% and C ranges from zero to one, F will range from 0% to 100%. Intelligibility is then predicted as a function of familiarity. In Section 6.3, step functions are used to predict intelligibility from familiarity.

6.2.2 Predicting interaction and contact

Contact can be measured by some of the techniques suggested in Section 6.1.2.3 and substituted straight into the above formulas; this is done in Section 6.3.5.1. A complementary method is to predict contact; this is done in Section 6.3.5.2. The prediction can be based on some of the factors underlying contact, such as distance between speech communities, population, relative distance from the center, and so on. The advantage of predicting from these factors is that they can be measured from maps and census data before going into the field. Predictions can also be based on overall attraction and motivation relations computed from the raw data (Section 6.1.2.3). When predictions are based on general patterns observed over the whole system, the estimated values may turn out to be better than the observed ones for at least two reasons: (1) the predicted values may afford a more refined measurement if the original scale had only two or three discrete levels, and (2) the predicted values may smooth over gross measurement errors in the raw data. The Santa Cruz Island data give evidence that contact predicted on the basis of overall attraction and motivation relations is a better predictor of intelligibility than the raw pairwise contact measurements (Section 6.3.5).

Models for predicting interaction and contact are not new to social science. They have been used by sociologists to explain human interaction for nearly one hundred years. Gerald Carrothers (1956) gives an extensive historical review of what are probably the most promising models, "gravitational" models. Quoting from Carrothers (1956:94):

> In general terms, the gravity concept of human interaction postulates that an attracting force of interaction between two areas of human activity is created by the population masses of the two areas, and a friction against interaction is caused by the intervening space over which the interaction must take place. That is, interaction between the two centers of population concentration varies directly with some function of the population size of the two centers and inversely with some function of the distance between them.

This is, of course, nothing more than an analogy to Newton's law of universal gravitation. The direct analogy to Newton's law is stated mathematically as follows (Carrothers 1959:95):

$$F_{ij} = \frac{P_i P_j}{D_{ij}^{2}}$$

where, F_{ij} = force of interaction between center _i_ and center _j_;

P_i, P_j = population of areas _i_ and _j_, respectively; and

D_{ij} = distance between center _i_ and center _j_.

Following the analogy from physics, the energy of interaction between the two communities would have given by multiplying the force times the distance. The result would have simply the distance, rather than the distance squared, in the denominator. For the distance term, not only geographic distance but also linguistic distance could be used to predict interaction. This would be based on an assumption that the greater the linguistic difference between dialects, the less the interaction that might be expected.

The above formula predicts reciprocal interaction. Of more interest to us than reciprocal interaction is one-way contact. In the sociological analogy to physics there is such a measure. It is termed "potential of population" (based on an analogy to potential energy) and was developed by Stewart (1941, Carrothers 1956:96). It suggests that the potential for interaction of an individual at i with the population of community j, will be greater as the population of j is greater, and will be less as the distance between i and j increases. Stated mathematically, the prediction of contact in these terms would be as follows:

$${}_iC_j = \frac{P_j}{D_{ij}}$$

where, ${}_iC_j$ = the potential contact of i with speech community j;

P_j = the population of community j; and

D_{ij} = the distance separating i and j.

In this formulation, the population of j serves essentially as a measure of the attraction of j. The assumption is, the larger j is, the more likely it is to attract contact. If we rewrite the formula to replace population with attraction, the result is a more general model which can have wider application in predicting language contact.

Another factor can be added to the model; this is the motivation of group i to have contact. The contact of i with the speech of community j does not depend solely on j's attraction, but also on i's motivation to interact. Some communities may be eager and outgoing; others may be cautious or reclusive.

Another refinement can be made to the model. When measuring the contact of one group with another, the distance between them can be measured relatively from the perspective of the group which is making the contact, rather than in absolute terms. The refined model for predicting contact is therefore,

$$ {}_iC_j = \frac{A_j M_i}{{}_iD_j} $$

where, ${}_iC_j$ = the contact of community i with the speech of community j;

A_j = the attraction of community j;

M_i = the motivation of community i; and

${}_iD_j$ = the distance from i to j as perceived by i.

There are many possibilities for specific factors to plug into the model. Distance could be geographic, or linguistic, or a combination of both. The following factors could be used to estimate the attraction of a group: population size, population density, nearness to the center of the dialect system, or the number of community facilities located within that group. Motivation could be estimated as the inverse of any of the above factors. That is, as the population of the group increases, we might expect its motivation to make contacts to decrease. As a group is nearer to the center of the dialect system, we might expect its motivation to make contacts to diminish, and so on. Another source of estimates for attraction and motivation are the sums of the rows and columns of the raw data matrices (Section 6.1.2.3). It was already suggested that the sums of the rows and colums reflect the relative motivation and attraction of the dialects within the whole dialect system.

Another possible perspective on attraction and motivation surfaced in the discussion of patterns of interaction (Section 6.1.2.2). There it was suggested that contact relations are the result of at least two factors, a need factor and an attitude factor. The degree to which i attracts j could be measured in terms of j's need to interact with i, and j's motivation to interact could be measured in terms of its attitude toward i.

The possibilities are numerous and at this point no dogmas concerning the best approach can be suggested. In the next section of this chapter some of the above proposals are tested against field data. These results serve to indicate the potential of the general model.

6.3 Explaining communication on Santa Cruz Island

The models developed in Section 6.2 are now used to explain communication on Santa Cruz Island in the Solomon Islands. The island is about 30 miles from end to end and has a total population of around 3000. In 1977 I conducted a dialect intelligibility survey there following the method described in Section 2.1. At the same time, Richard Buchan conducted a lexicostatistic survey. The results of both are reported in Simons 1977a.

The organization of this study is in seven parts. Section 6.3.1 reviews the data. Section 6.3.2 is an analysis which locates the center of the Santa Cruz dialect system. Section 6.3.3 reviews the statistical methods used to evaluate the models for explaining communication. Then in Sections 6.3.4 and 6.3.5 many different models for explaining communication are proposed and tested. Section 6.3.6 summarizes the successive refinements which were achieved at different stages in the modeling process. Finally, Section 6.3.7 draws conclusions, both of a specific nature for explaining communication on Santa Cruz Island and of a general nature for explaining communication elsewhere in the world.

6.3.1 The data

A complete description of the data for this study and all the data tables are found in Appendix 2.1. At this point I give only a brief review of what the data contain. The numbers in parentheses refer to the Appendix in which the details are found.

Intelligibility was tested in thirteen villages during the dialect survey. These thirteen villages represent the main dialects on Santa Cruz Island. In the first place, three items of information about each of the thirteen dialects are given: the villages which comprise them (2.1.1), the population of the dialects (2.1.2), and the density of population of the dialects (2.1.4). The remaining data consist of pairwise measurements of relationship between the dialects. These measurements include: the geographic distance between the dialects measured as traveling time (2.1.3), the lexical similarity between the dialects measured as cognate percentages (2.1.5), the lexical distance between the dialects measured as percentages of non-cognates (2.1.6), the intelligibility between dialects measured as described in Section 2.1 (2.1.7), local opinions about intelligibility (2.1.8), the contact of dialects through yearly church festivals (2.1.9), and the contact of dialects through marriage ties (2.1.10). Geographic and lexical distance are measured relatively as well as absolutely. In Appendix 2.1.11 a complete matrix of estimated intelligibility is given. The estimations are based on the final predicting model developed in Section 6.5.

6.3.2 The center of the Santa Cruz dialect system

The centrality of a dialect can be measured in several ways. It could be geographically central. It could be a center of population. It could be linguistically or culturally central. All these factors must be considered in defining a center. Ideally the evidence from many aspects of centrality will converge on a single answer. For Santa Cruz Island it does.

The center for the Santa Cruz dialect system is Mbanua (BAN). The evidence is summarized in the following list in which the first and second most central dialects are listed for each of the kinds of data presented in Appendix 2.1.

Data	First	Second
Population	BAN	LWO
Density	BAN	LWO
Geographic distance	BAN	NEP
Lexical distance	LWO	BAN
Intelligibility opinions	BAN, LWO, NEP, MAT, VEN (tie)	
Festival attendance	BAN	LWO
Marriage residence	BAN	LWO

The following criteria for defining centers were used: for population and density, the greatest population and density; for geographic and lexical distance, the lowest average distance; for opinions, festivals, and marriage, the greatest attraction.

The evidence clearly points to Mbanua as the central dialect. A further bit of evidence which has not yet been mentioned also concurs. This is the evidence of community facilities. Most of the dialects are self sufficient in terms of churches, stores, and primary schools.. The one facility which influences interaction on an island-wide scale is the administrative headquarters for the whole Eastern Outer Islands council area of the Solomon Islands. At this headquarters are located the only hospital, government offices, post office, police station, and airstrip for the island. The major ship wharf is there as well. This institutional center is located midway between Mbanua and Lwowa.

In Section 6.5 when models which predict contact on the basis of distance from the center are tested, the data used are the distances of the dialects from Mbanua. These are found in Appendix 2.1 by taking the BAN columns of Table 2.1 (geographic distance) and Table 2.4 (lexical distance).

6.3.3 The statistical method

Unfortunately, the statistical methods used in the analysis of lexical similarity and intelligibility (Chapter 5) are not applicable to these data. This is because intelligibility was measured on a discrete point scale, rather than on a continuous percentage scale. This results in two reasons why the least-squares methods of correlation and regression analysis used previously are not applicable to the current analysis: (1) the techniques are not appropriate for ordinal scale variables, and (2) the functions which predict intelligibility are step functions rather than linear functions.

Statisticians distinguish between ordinal level of measurement and interval level of measurement (Stevens 1946). When a variable is measured on an ordinal scale, each category has a unique position with respect to the other categories. That is, it is higher in value than some categories and lower in value than the rest. However, ordering is the only mathematical property of such a measurement scale; relative distance between categories is undefined. On the other hand, when a variable is measured on an interval scale an additional property characterizes the measurements. Not only are the categories ordered; the distances between the categories are defined in terms of fixed and equal units. Of the data described in Section 6.3.1, half of the variables (intelligibility, opinions, church festival attendance, and marriage ties) are measured on an ordinal scale. The techniques of correlation and regression analysis used in Chapter 5 require that the variables be measured on an interval scale.

In a linear function some amount of change in the independent variable always results in a proportional amount of change in the dependent variable. In a step function, change in the independent variable does not always result in change in the dependent variable. Rather, the dependent variable holds a constant value for specific ranges of the independent variable, and when the dependent variable changes it does so suddenly rather than gradually. When a step function is plotted on a graph, the appearance is that of a flight of stairs. Figure 6.3 illustrates a linear function and a step function. The techniques of correlation and regression analysis used in Chapter 5 apply when the underlying function is a linear one. In the current data, since intelligibility is measured on a four-point scale, functions which predict intelligibility will be step functions. Even the non-parametric correlation techniques (such as Spearman or Kendall rank-order correlations) which require no assumptions of linearity or interval scale variables do not handle step functions. With these techniques, a perfect step function yields less than perfect correlation.

For the two reasons just presented, it was necessary to use different methods of evaluating models for predicting intelligibility on Santa Cruz Island. In the previous chapter, the percentage of explained variation was the main measure used to evaluate the adequacy of a model. Here, the percentage of prediction accuracy is used. This percentage is based on the ratio of prediction accuracy. In the simplest case, this ratio would be obtained by dividing the number of correct predictions by the total number of predictions, which equals the number of cases.

However, in this formulation no account is made for how far off the incorrect predictions are. It would be good to distinguish between models with the same ratio of incorrect predictions but in which the errors are small in one model and large in another. The rationale here is that if administrative decisions had to be based on one of the two models, it would be better to use the one with the smaller errors. To do this, the number of correct predictions is decremented for predictions which are very wrong. In the Santa Cruz study understanding is at one of three levels: full intelligibility, partial intelligibility, or sporadic recognition. When a prediction is incorrect it can be off by one level or at most by two levels. When a prediction is off by two levels, one is subtracted from the number of correct predictions. Thus the ratio of prediction accuracy becomes,

$$\text{ratio of prediction accuracy} = \frac{\text{correct predictions} - \text{predictions off by two levels}}{\text{total predictions}}$$

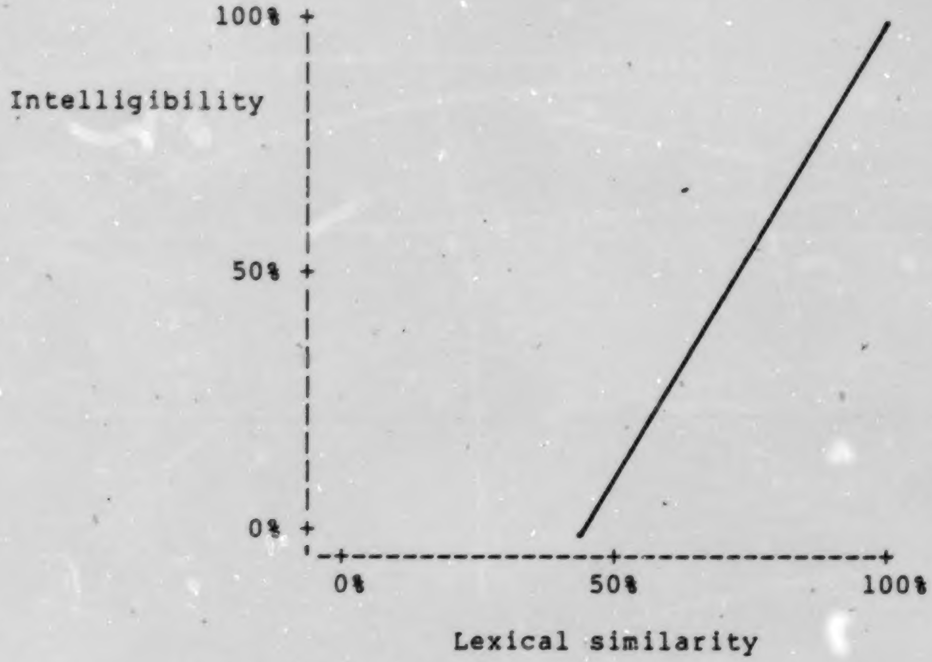
The same relationship can be formulated in another way using the concept of deviations. When a prediction is correct, the deviation from the measured value is zero; when it is off by one level, the deviation is one; when it is off by two levels, the deviation is two. The following formulation is therefore equivalent to the preceding one:

$$\text{ratio of prediction accuracy} = \frac{\text{total predictions} - \text{sum of deviations}}{\text{total predictions}}$$

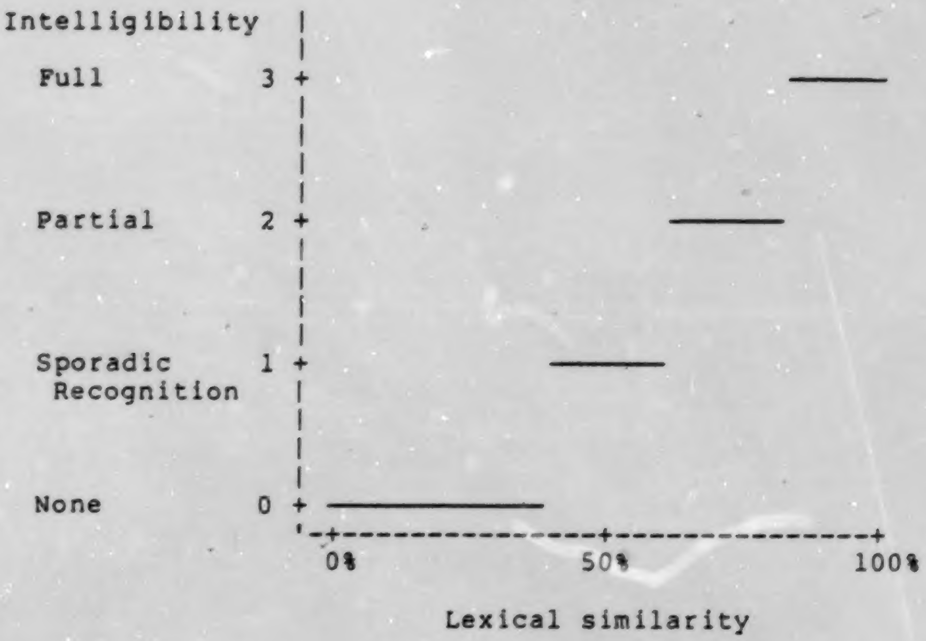
The percentage of prediction accuracy is then obtained by multiplying the ratio by one hundred.

Figure 6.3 A linear function and a step function

Linear function



Step function



In the previous chapter, the computational method of least squares was used to find the regression line and thus define the parameters of the model. In this analysis, since those techniques are not appropriate, the step function which best fits the data is found by inspection. First, a scattergram of the data is plotted. Then the cutoff points for the steps are located such that the percentage of prediction accuracy is maximized by minimizing the sum of the deviations. After fitting the step function, the sum of the deviations is totaled and the ratio and percentage of prediction accuracy are computed. The many scattergrams in Appendix 2.2 illustrate the technique.

Tests of significance are used to compare the accuracy of different models. The exact ratios of prediction accuracy are compared with a two-by-two chi-square test. Two types of tests are used. A two-tailed test is used to test the hypothesis that two ratios are unequal when there is no reason to suspect which one should be greater. A one-tailed test is used to test the hypothesis that one particular ratio is significantly greater than or less than another one. (One-tailed tests are always used unless specifically stated otherwise.) The result of the one- or two-tailed chi-square test is a significance level. The significance level is the probability that the two ratios are actually equal and that the observed difference could be due to chance alone. If this probability is very low, then we feel safe in accepting the hypothesis that the ratios are actually different (and that one is greater than the other in the case of a one-tailed test). Social scientists generally agree on the .05 level as being significant. When accepting a hypothesis at the .05 level, one is saying that there is no more than a one in twenty chance that it is wrong. In the discussions which follow, differences at the .05 level or better will be called "significant". Differences at the .01 level or better will be called "very significant". Differences at the .001 level or better will be called "highly significant". Occasionally differences just over the .05 level will be referred to as "nearly significant".

6.3.4 Single variable models

In Appendix 2.2 the scattergrams and the best fit step functions for single variable predictors of intelligibility are given. Six single variable predictors are tried: geographic distance, lexical similarity, opinions about intelligibility, church festival attendance, marriage ties, and predicted marriage residence. The accuracy of these single variable predictors is summarized in Figure 6.4. In this table, the percentage of prediction accuracy for the single variable predicates is given in the "With model" column. In parentheses, following the percentages, are the exact ratios of prediction accuracy. This format is followed in all remaining tables: percentages followed by the exact ratios. In the first column of numbers in the table, the percentage of prediction accuracy for the worst case model is given. This is the minimum percentage of accuracy that would be obtained if the relationship between intelligibility and the predictor were due to chance alone and not to any correlation between the two. In every case, the worst case occurs if the model predicts full intelligibility for all values of the independent value. This is because there are many more cases of full intelligibility than of the other two levels. Since the strategy in finding the best fit step function is to maximize prediction acccuracy, it can never be worse than what would be given by predicting only full intelligibility.

Note that in every case, the prediction accuracy for the model is greater than the worst case. The final column of the table gives significance levels for tests on the hypothesis that the single variable models are significantly better than chance associations (that is significantly better than the worst case). Three of

Figure 6.4 Single variable predictors of intelligibility

Predictor	Chance alone (worst case)	With model	Significance of improvement
Geographic distance	5[illegible] (44/78)	67% (52/78)	.09
Lexical similarity	56% (44/78)	77% (60/78)	.003
Opinions	56% (44/78)	77% (60/78)	.003
Festival attendance	61% (35/57)	79% (45/57)	.02
Marriage ties	56% (44/78)	58% (45/78)	.44
Marriage residence	56% (44/78)	62% (48/78)	.26

the predictors are 77% to 79% accurate and are definitely significant: lexical similarity, opinions, and festival attendance. Two are definitely not significant: marriage ties and marriage residence. The sixth predictor, geographic distance, is in a middle range where it is nearly significant. In the remaining figures in this chapter, significance with respect to the worst case is not computed in the tables because in every case the results are significant. An accuracy of 69% is significantly greater than the worst case of 56% at the .05 level (computed on ratios with a denominator of 78).

When the relations of lexical and geographic distance are made nonsymmetric by considering distance relative to the system, the result is a significant improvement in prediction accuracy over the symmetric measure of absolute distance in Figure 6.4. This is shown in Figure 6.5. The column of significance figures shows that the increase from 67% to 83% prediction accuracy for geographic distance is very significant. The increase from 77% to 86% for lexical distance is only nearly significant at the .07 level. However, the overall effect of measuring distance as relative rather than absolute (which is obtained by summing the results for geographic and lexical distance) is an increase from 72% to 85% which is highly significant.

Inspection of the scattergrams and step functions for relative lexical and geographic distance (see the last three scattergrams in Appendix 2.2) shows that relative geographic distance is a better predictor of intelligibility in the low intelligibility range while relative lexical distance is a better predictor in the high intelligibility range. That is, the greatest number of incorrect predictions for relative geographic distance are underestimates when measured intelligibility is "full intelligibility" and the greatest number of incorrect predictions for relative lexical distance are overestimates when measured intelligibility is "sporadic recognition". Since the strengths and weaknesses of the two different models are complementary, it follows that a combination of the two predicting variables might balance the weaknesses and yield a better prediction. This is indeed the case. An

Figure 6.5 Absolute, relative, and composite distance

	Absolute	Relative	Composite
Geographic distance	67% (52/78)	83% (65/78)	90% (70/78)
Lexical distance	77% (60/78)	86% (67/78)	90% (70/78)
Overall	72% (112/156)	85% (132/156)	90% (70/78)

Significance of:

	Relative > absolute	Composite > relative	Composite > absolute
Geographic distance	.008	.12	.0002
Lexical distance	.07	.23	.02
Overall	.003	.14	.001

optimal combination of the two relative measures of distance would be one which maximizes the prediction accuracy of the new composite variable. By iterating over various weightings of the two variables at steps of one-hundredth, it was found that the optimal combination is a combination consisting of 40% relative lexical distance and 60% relative geographic distance. That is,

Composite relative distance =
.4 X relative lexical distance +
.6 X relative geographic distance

Figure 6.5 shows that composite relative distance predicts intelligibility with an accuracy of 90%. This is an increase above 83% for relative geographic distance and 86% for relative lexical distance. However, tests of significance show that the size of these increases is not significant. The total improvement from the individual measures of absolute distance, however, to composite distance prove to be very significant for lexical distance and highly significant for geographic distance.

6.3.5 Complex models including linguistic similarity and contact

In the preceding section, simple models were considered in which intelligibility was viewed as a function of one factor, either linguistic similarity or contact. In this section, complex models in which intelligibility is viewed as a function of both similarity and contact are considered. The basic model has already been introduced in Section 6.2.1. It is as follows,

$$F = L + C(100-L)$$
$$I = f(F)$$

First, linguistic similarity and contact combine directly to predict the "linguistic familiarity", or F. The familiarity is a percentage estimate of what portion of the dialect of the speaker is familiar to the hearer, either through similarity or contact or both. Second, intelligibility is predicted from the familiarity. The function which maps familiarity onto intelligibility is a step function. When an actual model is specified it is necessary to make the function explicit by stating the ranges of F for each of the values of I.

Only one measure of linguistic similarity, L, is available in the present data. That is lexical similarity expressed as a percentage of cognates in basic vocabulary. This measure will be used for linguistic similarity in all formulations.

The object of the investigation in this section is to explore different variables which can be substituted for the contact factor, C, in the familiarity formula. First, measured values of contact are used. Then predicted values are tried. The predicted values are based on the attraction and motivation model presented in Section 6.2.2. Attraction and motivation are estimated first from measured contact, then from population relations, and finally from distance to the center of the dialect system.

In order to insure that the linguistic familiarity does not exceed 100%, the contact factor must be limited to a range of zero to one. For instance, measured contact through church festivals has the range zero to two. To adjust this variable for inclusion in the familiarity formula, the values need to be divided by two.

The situation for a variable like geographic distance is not as simple. The values of that variable measured in traveling time range from 5 minutes to 830 minutes. Furthermore, in making the adjustment the values must be inverted; that is, a high value of geographic distance implies low contact, and a low distance implies high contact.

The method used is this. First examine the distribution of values to determine the desired "minimum" and "maximum" values. These are not the true minimum and maximum; rather, they are the value which is to adjust to zero, and the one which is to adjust to one, respectively. For geographic distance, 830 is the minimum value. For the maximum value, 105 was selected on the following basis: on the island, it was observed that when neighboring dialect groups were within 105 minutes of traveling time, contact was always so great as to give the appearance of complete familiarity. The adjustment is made using this formula:

adjusted value = (original value - min) / (max - min)

Then if the adjusted value exceeds one, it is set to one. If it is less than zero, it is set to zero, unless negative values are used to reflect negative attitudes (they are not in this study). The minimum and maximum values for the scaling of all contact variables are reported in Appendix 2.3.

6.3.5.1 Measured contact

Five different kinds of measured contact are substituted into the familiarity formula to predict intelligibility. The scattergrams, best-fit step functions, and statistics for these five measures of contact are found in Appendix 2.4. The percentages of accuracy for the five predicting models are summarized in Figure 6.6.

In the top half of Figure 6.6, the hypothesis that predictions based on similarity and contact are more accurate than predictions based on contact alone is tested. The first column of numbers gives the prediction accuracies for models with contact alone; these are copied from Figure 6.4. The next column gives the prediction accuracies for complex models which combine the given contact factor with lexical similarity. The final column gives results of the significance tests on the hypothesis that the complex models are more accurate than the simple ones. In all five cases the complex model has the higher prediction accuracy. In three cases -- geographic distance, marriage ties, and marriage residence -- the improvement is significant. The overall effect, obtained by combining all the ratios in the columns, is a highly significant increase in prediction accuracy.

The bottom half of Figure 6.6 tests the hypothesis that the complex models combining similarity and contact give better predictions than the model based on similarity alone. The first column of numbers gives the prediction accuracy for the similarity model from Figure 6.4. The second column gives the prediction accuracies for the complex models. The final column gives results of the significance tests on the hypothesis that the complex models are more accurate than the similarity model. In only three of the five cases is there any increase in accuracy, and this is never significant. The overall effect, as well, shows no significant improvement of the complex models based on measured contact and similarity over the simple similarity model.

6.3.5.2 Predicted contact

Predicted contact is calculated for seven different factors. The first three are based on overall attraction and motivation measures for opinions about intelligibility, church festival attendance, and marriage residence. The remaining four are based on population, density of population, geographic distance from the center of the dialect system, and lexical distance from the center. The basic formula for predicting contact is the one developed in Section 6.2.2:

$$\text{Contact} = (\text{Attraction} \times \text{Motivation}) / \text{Distance}$$

Attraction is estimated by the overall attraction, the population or its density, or the inverse of the distance from the center. Motivation is estimated by the overall motivation, the inverse of population or density, or the distance from the center. Five different measurements of distance are used: absolute and relative geographic distance, absolute and relative lexical distance, and composite relative distance (six-tenths geographic and four-tenths lexical). The method in which each of these variables was scaled to a range of zero to one is given in Appendix 2.3. The distance measures were actually inverted and then scaled. In this way, predicted contact becomes the product of attraction, motivation, and inverted distance. Since the three component variables range from zero to one, the resulting predicted value

Figure 6.6 Measured contact

Contact variable	Contact Alone	With similarity	Significance of improvement
Geographic distance	67% (52/78)	78% (61/78)	.05
Opinions	77% (60/78)	83% (65/78)	.16
Festival attendance	79% (45/57)	84% (48/57)	.23
Marriage ties	58% (45/78)	74% (58/78)	.01
Marriage residence	62% (48/78)	77% (60/78)	.02
Overall	68% (250/369)	79% (292/369)	.0002

Contact variable	Similarity alone	With contact	Significance of improvement
Geographic distance	77% (60/78)	78% (61/78)	.43
Opinions	77% (60/78)	83% (65/78)	.16
Festival attendance	77% (60/78)	84% (48/57)	.15
Marriage ties	77% (60/78)	74% (58/78)	*
Marriage residence	77% (60/78)	77% (60/78)	.50
Overall	77% (60/78)	79% (292/369)	.33

* Since the second column is lower, the significance of the hypothesis that the second column is greater cannot be tested.

of contact is guaranteed to range from zero to one.

For each of the seven variables, eighteen different sets of contact predictions were made. These eighteen sets are organized into two intersecting dimensions containing three and six members. The first dimension represents the numerator of the contact formula. Three different numerators are tried: attraction alone, motivation alone, and attraction times motivation. The second dimension represents the denominator of the contact formula and six different denominators are tried. The first is a constant value of one which has the effect of leaving distance out of the formula. The other five are the five types of distance already mentioned. The results of the eighteen sets of predictions for the seven variables are given in full in Appendix 2.5.

The purpose of trying so many possible ways to predict contact is to test some hypotheses about which kinds of predictions are better and which are worse. The following hypotheses are tested:

(1) Relative distance predicts better than absolute distance;

(2) Relative distance predicts better than no distance;

(3) Composite distance predicts better than either measure of relative distance alone;

(4) Attraction and motivation together predict better than either alone;

(5) Predicted contact predicts better than measured contact;

(6) Contact predictions based on indirect measurements are equally as accurate as those based on direct contact measurements made in the field.

As is shown in the following paragraphs, hypotheses (1), (2), (5), and (6) can be accepted; (3) and (4) cannot.

The first hypothesis, that relative distance in the denominator of the contact formula is a better predictor than absolute distance, is tested in Figures 6.7 and 6.8. Figure 6.7 tests the hypothesis for geographic distance, while Figure 6.8 tests it for lexical distance. The two figures are otherwise completely parallel in their layout. The ratios on which the prediction accuracies are computed are the pooled results of the three different numerators for the contact equation: attraction alone, motivation alone, and attraction times motivation. Thus the ratios are the sums of the columns in the tables of results in Appendix 2.5.

In Figures 6.7 and 6.8, the first column of numbers gives the prediction accuracy for absolute distance models while the second column gives them for relative distance models. The last column gives the significance levels for accepting the hypothesis that the percentages in the second column are greater than those in the first. For both geographic and lexical distance, the percentage of accuracy for relative distance models is always greater than for absolute distance models. For models based on opinions and lexical distance from the center it is also always significantly so. For population, density, and marriage residence, however, it is not. For festival attendance and geographic distance from the center, the increase is nearly significant. For both geographic and lexical distance the overall increase in prediction accuracy (obtained by summing the

Figure 6.7 Absolute versus relative geographic distance

A and M predictor	Absolute distance	Relative distance	Significance of improvement
Opinions	78% (182/234)	89% (208/234)	.0008
Festival attendance	76% (130/171)	82% (141/171)	.07
Marriage residence	75% (175/234)	80% (187/234)	.09
Population	78% (183/234)	81% (190/234)	.21
Density	79% (186/234)	81% (190/234)	.32
Geographic center	80% (187/234)	85% (200/234)	.06
Lexical center	81% (189/234)	87% (204/234)	.03
Overall	78% (1232/1575)	84% (1320/1575)	.00003

Figure 6.8 Absolute versus relative lexical distance

A and M predictor	Absolute distance	Relative distance	Significance of improvement
Opinions	80% (188/234)	90% (211/234)	.001
Festival attendance	78% (134/171)	84% (144/171)	.08
Marriage residence	78% (182/234)	82% (191/234)	.15
Population	78% (182/234)	81% (189/234)	.21
Density	79% (186/234)	83% (194/234)	.17
Geographic center	80% (188/234)	87% (203/234)	.03
Lexical center	82% (193/234)	88% (206/234)	.05
Overall	80% (1253/1575)	85% (1338/1575)	.00004

columns) is highly significant. Thus I accept the hypothesis that relative distance in the denominator of the contact formula is a better predictor than absolute distance.

The second hypothesis, that relative distance in the denominator of the contact formula is a better predictor than no distance at all, is tested in Figure 6.9. In that figure the test against relative lexical distance is shown. A test against relative geographic distance would have very similar results. The table parallels Figures 6.7 and 6.8 in its construction. The results in the significance column show that in only two cases was the increase not significant, but only nearly significant. The overall increase in prediction accuracy is highly significant and I therefore accept the hypothesis.

Figure 6.9 No distance versus relative lexical distance

A and M predictor	No distance	Relative distance	Significance of improvement
Opinions	82% (193/234)	90% (211/234)	.008
Festival attendance	76% (130/171)	84% (144/171)	.03
Marriage residence	70% (164/234)	82% (191/234)	.002
Population	73% (170/234)	81% (189/234)	.02
Density	74% (173/234)	83% (194/234)	.009
Geographic center	82% (193/234)	87% (203/234)	.10
Lexical center	83% (194/234)	88% (206/234)	.06
Overall	77% (1217/1575)	85% (1338/1575)	.0000001

The third hypothesis, that composite relative distance in the denominator of the contact formula is a better predictor than either relative geographic or lexical distance alone, is tested in Figure 6.10. The three columns of numbers represent the three different measures of distance: relative geographic, relative lexical, and composite relative. The ratios are based on a pooling of the results for the three different numerators as before. Within the table there are no significant differences. Below the table, the results of significance tests on the overall trends are given. The overall trend is not even in the direction of the original hypothesis. That is, relative lexical distance turns out to have a higher overall prediction accuracy than composite distance. Therefore, two-tailed tests of significance are made. These test the hypothesis that the prediction accuracies are at all different, without specifying the direction of the difference. In the best case, we would be wrong 38% of the time if we suggested that relative geographic and relative lexical distance gave different results. There is no basis for accepting a hypothesis that any one of these three measures of distance gives significantly better results than the other two.

The fourth hypothesis, that models with attraction times motivation in the numerator of the contact formula are better predictors than models with just attraction or just motivation, is tested in Figure 6.11. The ratios in the body of the table are based on a pooling of three models: with relative geographic distance in the denominator, with relative lexical distance, and with relative composite distance. These are the best models: the ones with absolute distance and no distance are not included since they have been shown to be significantly less accurate. In the overall trend, the attraction-times-motivation models prove to have the highest ratio of prediction accuracy. The tests of significance show that the increase over models of attraction alone is significant, but only at the .05 level. The increase over motivation alone models is not significant. Looking within the body of the table, it is apparent that there is no strong trend. In one case the attraction alone model is the best and in two cases the motivation alone models are the best. In only four pairs of models is one significantly less than the other. Attraction alone for opinions is significantly less than the other two opinion models, and motivation alone for festival attendance is significantly less than the other two. The results are thus largely inconclusive. In is not possible to conclude that any particular combination of attraction or motivation or both yields better prediction accuracy in general.

The fifth hypothesis, that predicted contact is a better predictor than measured contact, is tested in Figure 6.12. There are only three contact factors for which the accuracy of measured and predicted values can be compared. The percentages and ratios of prediction accuracy for measured contact are copied from Figure 6.6. The figures for predicted contact are taken from Appendix 2.5. For each of the three contact factors, the most accurate model is chosen to fill in the predicted contact column. Comparison of the columns shows that in all three cases the accuracy with predicted contact is higher than with measured contact. In the case of opinions it is significantly higher; in the case of festival attendance it is not; in the case of marriage residence it is nearly so. The overall trend shows that predicted contact raises the accuracy from 81% to 89% which is a significant increase at the .02 level.

The sixth hypothesis, that contact predictions based on indirect measurements are equally as accurate as those based on direct measurements in the field, is tested in Figure 6.13. For each of the seven contact factors, the best model from the tables in Appendix 2.5 is chosen to represent it. The seven factors are grouped into three categories and overall accuracy for each category is computed. The overall accuracy forr direct predictions based on measured contact is 89%, for indirect predictions based on population statistics it is 84%, and for indirect predictions based on distance from the center of the dialect system it is 88%.

The percentages for overall direct predictions and overall distance from center predictions are nearly equal. A two-tailed test is used to test the hypothesis that they are equal. The result shows that the probability that they are the same is 98%. The conclusion, therefore, is that the models using indirect contact predictions based on distance from the center of the dialect system are equally as accurate as the models which use direct contact predictions based on field measurements of contact. The prediction accuracy for predictions based on population statistics is less than for the other two categories. Thus tests were

Figure 6.10 Composite versus single relative distance

	Geographic	Lexical	Composite
Opinions	89% (208/234)	90% (211/234)	90% (210/234)
Festival attendance	82% (141/171)	84% (144/171)	83% (142/171)
Marriage residence	80% (187/234)	82% (191/234)	80% (188/234)
Population	81% (190/234)	81% (189/234)	82% (191/234)
Density	81% (190/234)	83% (194/234)	81% (190/234)
Geographical center	85% (200/234)	87% (203/234)	86% (201/234)
Lexical center	87% (204/234)	88% (206/234)	87% (204/234)
Overall	84% (1320/1575)	85% (1338/1575)	84% (1326/1575)

Significance tests on overall trends (two-tailed tests):

Geographic ≠ Lexical	.38
Composite ≠ Lexical	.56
Geographic ≠ Composite	.78

Figure 6.11 Attraction and motivation

	Attraction alone	Motivation alone	Attraction and motivation
Opinions	86% (202/234)	91% (213/234)	91% (214/234)
Festival attendance	86% (147/171)	79% (135/171)	85% (145/171)
Marriage residence	79% (186/234)	83% (194/234)	79% (186/234)
Population	81% (189/234)	82% (191/234)	83% (195/234)
Density	81% (189/234)	81% (190/234)	83% (195/234)
Geographic center	85% (199/234)	87% (203/234)	86% (202/234)
Lexical center	85% (199/234)	88% (206/234)	89% (209/234)
Overall	83% (1311/1575)	85% (1332/1575)	85% (1346/1575)

Significance tests for overall trends:

Motivation	> Attraction	.15
A and M	> Motivation	.24
A and M	> Attraction	.05

Figure 6.12 Measured contact versus predicted contact

	Measured	Predicted	Significance of improvement
Opinions	83% (65/78)	92% (72/78)	.04
Festivals attendance	84% (48/57)	88% (50/57)	.29
Marriage residence	77% (60/78)	86% (67/78)	.07
Overall	81% (173/213)	89% (189/213)	.02

Figure 6.13 Direct predictions versus indirect predictions

Direct predictions based on measured contact:

Opinions	92% (72/78)
Festival attendance	88% (50/57)
Marriage residence	86% (67/78)
Overall	89% (189/213)

Indirect predictions based on population statistics:

Population	83% (65/78)
Density of population	85% (66/78)
Overall	84% (131/156)

Indirect predictions based on distance from center of dialect system:

Geographic distance from center	87% (68/78)
Lexical distance from center	90% (70/78)
Overall	88% (138/156)

Significance tests:

Overall direct ≠ overall distance from center	.98 (two-tailed)
Overall direct > overall population	.09
Opinions > overall population	.04
Overall distance from center > overall population	.12

made to see if they are significantly less accurate. The significance levels for the tests are .09 and .12 so it is not possible to conclude that the indirect predictions based on population statistics are significantly less accurate than predictions based on measured contact or distance from the center. However, a test of the best single model, the opinions model at 92% accuracy, against the overall accuracy of population statistics shows that they are significantly less accurate than the best model.

Thus far, no attempt has been made to combine different contact factors in predicting intelligibility. Obviously, contact has many facets and predictions which simultaneously account for many aspects of contact, rather than considering them only one at a time, should be better predictions. A simple way to use step functions in making predictions which combine factors is to select an odd number of factors and for each case to predict the level of intelligibility indicated by the majority of the factors considered individually. (This method has the weakness that it does not consider the possible interaction of factors.) This is done with three factors -- composite relative distance alone, predicted contact based on opinions, and predicted contact based on lexical distance from the center -- for the Santa Cruz data. In most cases, all three predictions agree. Where they do not, the level estimated by two out of the three factors is taken as predicted intelligibility. The resulting predictions are 95% accurate. More details on the method used and a complete matrix of estimated intelligibility based on this combined approach are given in Appendix 2.1.11.

6.3.6 Summary of refinements

A summary of the refinements which have been made in Sections 6.3.4 and 6.3.5 to models for explaining communication is given in Figure 6.14. In that chart the percentage of prediction accuracy for different types of models are given. The arrows show the directions of refinement as new factors are combined to improve the accuracy of the models. The numbers on the arrows are the significance level for a test of the hypothesis that the model at the head of the arrow is more accurate than the one at the tail.

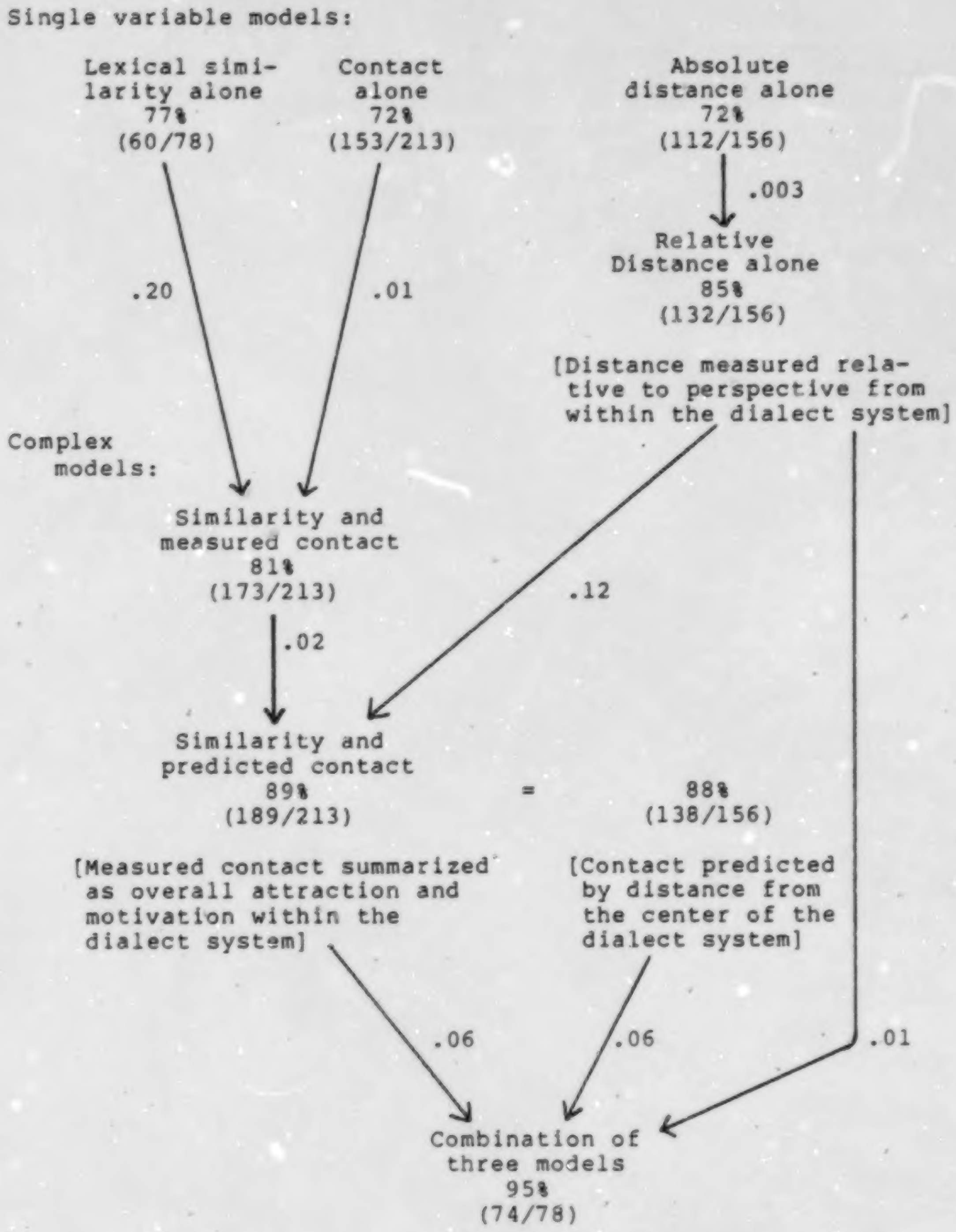
The initial models are single variable models. A model which explains intelligibility as a function of lexical similarity alone is 77% accurate (from Figure 6.4). Overall, models which explain intelligiblity as a function of some single factor of contact are 72% accurate. This estimate is based on a pooling of results from opinions, festival attendance, and marriage residence from Figure 6.4; only these three are considered in order to maintain comparability at further steps in the development. These degrees of accuracy are significantly greater than what is possible by chance alone. 77% accuracy for lexical similarity is greater than 56% for the worst case (44/78) at .003 significance; 72% accuracy for contact is greater than 58% for the worst case (123/213) at .001 significance.

When lexical similarity and contact are combined in a more complex model to explain intelligibility, the degree of accuracy increases to 81% (Figure 6.6). This is greater than the accuracy for similarity alone by a confidence level of .20, and greater than the accuracy for contact alone by a confidence level of .01. The model used to predict intelligibility is,

$$I = f(L + C(100 - L))$$

where $\underline{I}$ represents the level of intelligibility, $\underline{L}$ represents the percentage of lexical similarity, and $\underline{C}$ represents the degree of contact. The contact measures

Figure 6.14 Summary of successive refinements to models for explaining communication


Single variable models:
Lexical similarity alone 77% (60/78)
Contact alone 72% (153/213)
Absolute distance alone 72% (112/156)
.003
Relative Distance alone 85% (132/156)
.20
.01
[Distance measured relative to perspective from within the dialect system]
Complex models:
Similarity and measured contact 81% (173/213)
.12
.02
Similarity and predicted contact 89% (189/213)
=
88% (138/156)
[Measured contact summarized as overall attraction and motivation within the dialect system]
[Contact predicted by distance from the center of the dialect system]
.06
.06
.01
Combination of three models 95% (74/78)

used as predictors in the single variable models were scaled to a range of zero to one and then plugged directly into the prediction formula.

Another kind of single variable model tested was a model based on absolute geographic or lexical distance between dialects. Those models were 72% accurate overall (Figure 6.5). When the measure of distance is refined by making it relative to the perspective from within the dialect system, then the accuracy increases to 85% (Figure 6.5). This increase is significant at the .003 level.

A further refinement to models for explaining communication is to predict contact rather than to measure it. The resulting models combine the threee single variable predictors already discussed -- lexical similarity, measured contact, and distance -- and are 89% accurate overall (Figure 6.12). This accuracy is greater than the complex models using measured contact at .02 significance, and is greater than the accuracy for models with distance alone at the .12 level. The formula used to predict contact is,

$$C = AM/D$$

where C represents the hearer's contact with the speech of the speaker, A represents the attraction of the speaker's group, M represents the motivation of the hearer's group to have contact, and D represents the distance from the hearer's group to the speaker's. Relative distance is used for the distance term. Attraction and motivation are estimated by the overall attraction and motivation of dialect groups as indicated in the measured contact data.

Another refinement simplifies the task of data collection with no significant loss in accuracy of the model. The same formula is used as in the predicted contact models in the previous paragraph. The difference is that attraction and motivation are estimated by the distance of the dialects from the center of the dialect system. This kind of estimation does not require the collection of pairwise contact measurements in the field as the method in the previous paragraph does. The overall accuracy of these models which predict contact by distance from the center is 88%. This is not significantly different from the accuracy of the estimates based on pairwise field measurements (Figure 6.13).

A final refinement is to combine different contact factors in predicting intelligibility. Since contact involves many factors, to consider different aspects in combination gives better predictions than to consider any one aspect by itself. When this is done for the Santa Cruz data, the resulting predictions are 95% accurate.

6.3.7 Conclusions

6.3.7.1 Explaining communication on Santa Cruz Island

Certain conclusions concerning what explains communication on Santa Cruz Island can be made from this study. First of all, the lexical similarity between dialects is an important factor. By itself it can correctly account for the level of intelligibility in 77% of the cases. Secondly, contact between dialects, in combination with similarity, accounts for half of the remaining incorrect cases in general (88% accuracy, Figure 6.14) and over three-quarters of the remaining incorrect cases in the final combination model (95% accuracy, Figure 6.14).

The results give some indication of what aspects of contact are most significant in explaining communication on Santa Cruz Island. Marriage ties and predicted marriage residence turn out to be the least effective explainers of communication. The frequency with which dialects have contact at yearly church festivals proves to be more effective. The significant increase in accuracy of relative distance models over absolute distance models indicates that a Santa Cruz speech community's motivation to get out and travel long distances to make contact increases as its distance (both geographic and linguistic) away from the other speech communities increases. The success of the models which predict contact on the basis of distance from the center of the dialect system suggests that Santa Cruz Island is indeed a centered dialect system. The interpretation of those models is that the nearer a dialect is to the center, the more likely it is to attract contact and the less motivated it is to go out and make contact. Conversely, the further a dialect is from the center, the more motivated it is likely to be to go out and make contact and the less likely it is to attract contact. The result is a general directing of contact relations in toward the center.

6.3.7.2 Explaining communication elsewhere

This study of models for explaining communication on Santa Cruz Island suggests at least three conclusions which could have general application: (1) the value of local opinions about intelligibility, (2) the systemic nature of dialect relations, and (3) the potential of the modeling method.

The single best predictor of intelligibility turned out to be local opinions about intelligibility. Many investigators have stayed clear of informant opinions because they are so open to a subjective element. At first glance the same conclusion might be reached for Santa Cruz Island. Opinions alone are only 77% accurate at explaining intelligibility (Figure 6.4). I would attribute the errors not so much to errors in the informants' judgments as to the clumsiness of the method I used to measure opinions. I found it possible to elicit responses at only three degrees of understanding with any consistency -- understand all, some, or none. Of course, degree of understanding covers a continuous range. However, it was possible to reconstruct predictions of a continuous nature from the original opinions. This was possible when the whole island was viewed as a dialect system and all of the opinions about a dialect were viewed as saying something about its ability to attract communication and all of the opinions given by a dialect were viewed as saying something about its motivation to communicate. Predictions based on these refinements of the original opinions are 92% accurate (Figure 6.13); this 15% improvement is significant at a .004 level. Refinements such as these may in general increase the value of local opinions concerning intelligibility.

he key to explaining communication on Santa Cruz Island is viewing the dialects as comprising a dialect system. In the summary of the results given in Figure 6.14, comments in square brackets are placed at three spots in which the system viewpoint plays a significant role. These are as follows: (1) The measurement of distance as relative to perspective from within the system rather than in absolute units (see Section 6.1.3) proved to significantly increase the accuracy of single variable models (Figure 6.5) and complex models (Figures 6.7 and 6.8). (2) The pairwise measurements of contact taken in the field -- opinions, festival attendance, marriage ties -- were made in terms of three or four point discrete scales. The degree of discrimination possible in predictions based directly on these measurements is therefore not very great. However, when the individual pairwise measurements are viewed as being interrelated within a dialect system, all of the measurements concerning individual dialects can be summed to compute that dialect's overall attraction and motivation within the system.

Predictions based on these attraction and motivation estimates are significantly better than those based on the original measurements (Figure 6.12). Relations based on the whole system may not only offer a more discriminating measure than the original data, they may also serve to compensate for measurement errors in the original data. (3) One of the characteristics of a dialect system is that general patterns of interaction can be explained in terms of common relations to a center (Section 6.1.4). It was found that predictions based on distance from the dialect center were as accurate as the predictions based on overall attraction and motivation relations found in the original contact measurements (Figure 6.13). A model which explains communication in terms of relations to a center is one order of magnitude simpler than one which relies on all of the pairwise relations in the system. That is, if there are n dialects, the centered model requires n raw data measurements (the distance of each dialect from the center) while the other model uses n-squared raw data measurements (to fill in a square matrix of pairwise contact measurements). This increased simplicity is significant not only because it reduces the complexity of data and model description, but it could also reduce the complexity of data collection.

A final conclusion regards the potential of this modeling appraoch. Three models based on a single contact factor were 90% accurate or better: one based on composite relative distance alone and two complex models with predicted contact based on opinions or lexical distance from the center. When three factors were combined by taking the level of intelligibility predicted by at least two of the three factors, 95% accuracy was achieved. When the accuracy of intelligibility predictions based on linguistic similarity and contact begins to exceed 90%, one begins to wonder if the measured intelligibility data themselves are 90% accurate. When predictions are that accurate they become useful indices by which to evaluate the measured intelligibility. Of course, in this study, prediction accuracy was measured by comparing the predicted values to the measured values. Thus we could not have done anything without the intelligibility measurements. However, this is a pilot study. As we come to better understand the workings of dialect systems through further study, it may become possible to one day predict intelligibility, even without first measuring it, with an accuracy greater than that with which we could have measured it.

APPENDIX 1

COMPLETE DATA FOR THE STUDY OF LEXICAL SIMILARITY AND INTELLIGIBILITY

1.1 Sources of data

For each set of data six items of information are given: the source of the intelligibility data, a brief note on the method of intelligibility testing, the type of adjustment used to control for intelligibility measurement error (Section 5.2.4), the source of the cognate percentages, the type of word list used, and the correspondence of three letter mnemonic codes to village or dialect names. The ten studies are considered in alphabetical order of the name by which they are referenced.

(1) Biliau - Biliau is spoken in the Madang Province of Papua New Guinea. The dialect survey was conducted by myself and my wife, Linda, in 1976. The intelligibility testing followed the method of Casad with two exceptions: the questions were asked in the trade language and tests were administered to groups as well as to individuals. The raw intelligibility scores are adjusted by the hometown method. The word list used was the Swadesh 100-word list. The correspondence of mnemonic codes to village names is:

BIL = Biliau
YAM = Yamai
SUI = Suit

Unfortunately, the data here represent only the results of a pilot study; sickness prevented the completion of the full survey. Neither the data nor the results have been published elsewhere.

(2) Buang - Buang is spoken in the Morobe Province of Papua New Guinea. The survey of Buang dialects was conducted by Gillian Sankoff between 1966 and 1968. The intelligibility and lexicostatistic data are taken from Sankoff 1969. The approach to intelligibility testing is similar to that of Casad though not as exact. Subjects listened to a test tape and then answered three questions about it in order to judge comprehension of events in the story (see Section 2.2.4 for a fuller description). A proportional adjustment for subjects is used to adjust raw intelligibility scores. The test list used for lexicostatic comparison was a 162 item list comprising the Swadesh 100-word list plus a number of cultural items specific to New Guinea. The correspondence of mnemonic codes to village names is:

BUW = Buweyew
MNB = Mambump
WIN = Wins
CHI = Chimbuluk
PAP = Papekene
MNG = Mangga
KWA = Kwasang

Although the intelligibility tests were administered in seven different villages, only three dialects were used in test tapes -- MMB, CHI, and MNG. Thus for four of the villages there are no proper hometown scores. For these four the hometown score for one of the other three villages is used as an estimate. The one used is the one in the same dialect group, according to Sankoff's grouping into three dialects. Thus the hometown score for MMB serves also for BUW, the hometown score for CHI serves also for WIN and PAP, and the hometown score for MNG serves also for KWA.

(3) Ethiopia - The data from Ethiopia come from the intelligibility survey of the Sidamo languages conducted by Marvin Bender and Robert Cooper (1971). The test method consisted of playing the test text, then having the subjects (who were school children) answer multiple choice questions with four possible responses about the contents of the story. These tests were written, and were conducted in the national language. The method is described in more detail in Section 2.2.4.

Since the tests consisted of choosing the correct one out of four possible answers, it is possible that a group of subjects with no knowledge of a language could score 25% correct simply by chance. Therefore, when subjects scored less than 25%, it can be assumed that there was no comprehension. Thus the raw scores must first be adjusted to remove the chance element. This is done by recomputing the scores as the percentage of correct responses above the chance level. The score for correct responses above the chance level is given by subtracting 25% from the raw score or by 0%, whichever is greater. The total possible above the chance level is given by subtracting 25% from 100%, or 75%. The percentage of intelligibility adjusted for chance is obtained by dividing the correct by the total possible and multiplying by 100. That is,

$$\text{adjusted for chance} = \max(\text{raw} - 25, 0) / 75 \times 100$$

The "raw" scores reported in Appendix 1.2 have already been adjusted in this way. Bender and Cooper made no such adjustment; the technique was suggested by Ladefoged, Glick, and Criper (1972:68). The adjusted scores for this set of data are further adjustments on these raw scores. For this, the hometown adjustment was used.

The cognate percentages to accompany the intelligibility scores are found in another source, Bender (1971). The word list used was the Swadesh 100-word list with modifications dictated by experience in the Ethiopian field. Correspondence of mnemonic codes to dialect names is:

ALA = Alaba
KEM = Kembata
HAD = Hadiyya
SID = Sidamo
DER = Derasa
BUR = Burji

(4) Iroquois - The intelligibility survey among the Iroquois languages, northeastern United States, was conducted by Hickerson, Turner, and Hickerson (1952). Their method was basically one of text translation and the study has been described in more detail in Chapter 2, Section 2.2.1. A proportional adjustment for subjects was used to compute the adjusted intelligibility scores. The cognate percentages are taken from Floyd Lounsbury (1961). The word list used for comparison was the Swadesh 200-word list. Hickerson, Turner, and Hickerson tested intelligibility among six dialects, resulting in 36 pairwise measurements. The

lexicostatistic comparisons by Lounsbury involve only four of those six dialects. Within the set of lexicostatistic comparisons, the percentage for one pair of languages (Tuscarora with Cayuga) is missing. As a result, corresponding lexical data were found for only fourteen of the thirty-six intelligibility measurements. Only these fourteen data points are included in the sample. The correspondence of mnemonic names to dialect codes is:

SEN = Seneca
CAY = Cayuga
ONE = Oneida
TUS = Tuscarora

(5) *Mazatec* - The intelligibility survey among the Mazatec dialects of Mexico was carried out by Paul Kirk. The results of the survey were first published by Kirk (1970) and then reproduced by Casad (1974:34-35, 47-49). The method used in the testing was the Casad method (Section 2.2.3). The raw intelligibility scores are adjusted by the hometown method. The lexicostatistic comparison of the Mazatec dialects was done by Sarah Gudschinsky (1955). The Swadesh 200-word list was used for the comparison. The correspondence of mnemonic codes to village names is:

HUA = Huautla de Jimenez
MAT = San Mateo
MIG = San Miguel
IXC = Ixcatlan
SOY = Soyaltepec
JAL = Jalapa de Diaz

(6) *Polynesia* - Intelligibility among the Polynesian languages and dialects was tested by Jack Ward (1962). The method of testing used was a sentence translation test. A constant adjustment for subjects was used to adjust raw intelligibility scores. The lexicostatistic comparisons were performed by Samuel Elbert (1953). The comparisons are based on Swadesh's early basic vocabulary of 165 words which Elbert expanded to 202 words. The correspondence of mnemonic codes to language and dialect names is:

EAS = Easter Island
HAW = Hawaiian
KAP = Kapingamarangi
MAN = Mangareva
MAO = New Zealand Maori
MAR = Marquesas
RAR = Rarotonga
SAM = Samoa
TAH = Tahiti
TON = Tonga
TUA = Tuamotu
UVE = Uvea

(7) *Siouan* - The intelligibility survey among the Siouan languages of the Great Plains area of the United States and Canada was conducted by Warren Harbeck and Raymond Gordon in 1968. The results of the survey are reported in an unpublished paper (Harbeck ms [1969]). The method used for the testing was one of text translation. The tests were scored in two ways: the first computed the accuracy of an item by item translation and the second measured general comprehension by

checking for the presence of ten key pieces of information in the translation. The raw intelligibility scores used in this study are the average of the results of the two different scoring procedures. The raw intelligibility scores are adjusted by the hometown method. The cognate percentages are from the same source and are based on the Swadesh 100-word list. The correspondence of mnemonic codes to dialect names is:

STO = Stoney
ASS = Assiniboine
MAN = Manitoba variety of Dakota-Nakota
NDK = North Dakota variety of Dakota-Nakota
LAK = Lakota

(8) Trique - The intelligibility survey of the Trique language area of Mexico was carried out by Eugene Casad in 1970. The results of the survey are reported in his manual on dialect intelligibility testing (1974:78-81, 191-192). The method of testing was the question approach described in detail in the manual (Section 2.2.3). The raw intelligibility scores are adjusted by the hometown method. The lexicostatistic comparison was based on the Swadesh 100-word list. The correspondence of mnemonic codes to village names is:

MIG = San Miguel
ITU = Itunyoso
LAG = Laguna
CHI = Chicahuaxtla
SAB = Sabana

(9) Uganda - These data are the results of intelligibility tests conducted with speakers of two Bantu languages in Uganda by Peter Ladefoged (1968). The intelligibility test results were extracted from page 67 of Ladefoged, Glick, and Criper (1972), and the cognate percentages are extracted from page 71. The tests were administered to literate school children. A short story from another language was played and the listeners were asked what it was about. They were presented with three possible answers which were also written in a test booklet and were asked to write down the number of the appropriate response. The raw scores reported in Appendix 1.2 are adjusted to acccunt for the element of chance as already described in the description of the Ethiopian data, the only difference being that here the chance level is computed at 33.3%. The adjusted scores reported in Appendix 1.2 have undergone a further proportional adjustment for subjects. The word list used for lexicostatistic comparison was a list designed especially for the Ugandan survey. In setting up the new list the guiding principle was not to use basic vocabulary which is supposedly more resistant to change, but to use meanings which elicited reliable answers and which were valid indicators of the communicative possibilities of the language as a whole (Ladefoged and others 1972:54-55). The correspondence of mnemonic codes to language names is:

LUG = Luganda
RUN = Runyankore
RUT = Rutooro
RUK = Rukiga
LUM = Lumasaba
LUS = Lusoga

(10) Yuman - The intelligibility survey among the Yuman langauges of the southwestern United States was conducted by Bruce Biggs (1957). The method was basically a text translation approach and is described in detail in Chapter 2, Section 2.2.1. The cognate percentages which correspond to the intelligibility percentages are reported by Biggs. They are taken from Werner Winter (1957). A 100-word list was used. The word list items "were chosen at random, though with considerable emphasis on words from Swadesh lists" (Winter 1957:19). The correspondence of mnemonic codes to dialect names is:

MAR = Maricopa
WAL = Walapai
YAV = Yavapai
MOH = Mohave
HAV = Havasupai

1.2 Complete listing of data

The following pages are a complete listing of the data, presented study by study. The data are presented in eight columns. They are, in order: (1) "HEAR", the hearers, the mnemonic code of the village or dialect taking the intelligibility test; (2) "SPKR", the speaker, the mnemonic code of the village or dialect which is speaking on the test tape; (3) "LEX", the percentage of lexical cognates; (4) "INT RAW", the raw percentage of intelligibility; (5) "INT ADJ", the adjusted percentage of intelligibility (for each set, the method of adjusting is described in Appendix 1.1); (6) "EXCLUDE", an "X" is listed if this case is excluded due to nonsymmetric intelligibility attributed to social factors; (7) "SUBJ", the hometown score for the group of subjects (used in adjusting raw intelligibility); (8) "TEST", the hometown score for the test which is being administered (used in adjusting raw intelligibility).

(1) Biliau

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
BIL	BIL	100	92	100.0		92	92
BIL	YAM	98	90	90.0		92	95
BIL	SUI	82	88	88.0		92	95
YAM	BIL	98	100	100.0	X	95	92
YAM	YAM	100	95	100.0		95	95
YAM	SUI	81	90	90.0	X	95	95
SUI	BIL	82	100	100.0	X	95	92
SUI	YAM	81	80	80.0		95	95
SUI	SUI	100	95	100.0		95	95

(2) Buang

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
BUW	MMB	88	67	91.8		73	73
BUW	CHI	78	44	60.3		73	67
BUW	MNG	65	14	19.2		73	68
MMB	MMB	100	73	100.0		73	73
MMB	CHI	83	67	91.8	X	73	67
MMB	MNG	61	25	34.2		73	68
WIN	MMB	83	73	100.0	X	67	73
WIN	CHI	88	60	89.6		67	67
WIN	MNG	67	23	34.3		67	68
CHI	MMB	83	43	64.2		67	73
CHI	CHI	100	67	100.0		67	67
CHI	MNG	66	52	77.6	X	67	68
PAP	MMB	80	53	79.1	X	67	73
PAP	CHI	93	53	79.1		67	67
PAP	MNG	69	53	79.1		67	68
MNG	MMB	61	43	63.2	X	68	73
MNG	CHI	66	43	63.2		68	67
MNG	MNG	100	68	100.0		68	68
KWA	MMB	60	50	73.5	X	68	73
KWA	CHI	66	52	76.5		68	67
KWA	MNG	94	57	83.8		68	68

(3) Ethiopia

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
ALA	ALA	100	91	100.0		91	91
ALA	KEM	81	95	94.7	X	91	99
ALA	HAD	54	61	61.3		91	89
ALA	SID	64	28	28.0		91	95
ALA	DER	49	16	16.0		91	81
ALA	BUR	40	13	13.3		91	91
KEM	ALA	81	79	78.7		99	91
KEM	KEM	100	99	100.0		99	99
KEM	HAD	56	49	49.3		99	89
KEM	SID	62	23	22.7		99	95
KEM	DER	49	9	9.3		99	81
KEM	BUR	39	24	24.0		99	91
HAD	ALA	54	67	66.7		89	91
HAD	KEM	56	65	65.3	X	89	99
HAD	HAD	100	89	100.0		89	89
HAD	SID	53	25	25.3		89	95
HAD	DER	42	33	33.3	X	89	81
HAD	BUR	38	20	20.0		89	91
SID	ALA	64	40	40.0	X	95	91
SID	KEM	62	16	16.0		95	99
SID	HAD	53	25	25.3		95	89
SID	SID	100	95	100.0		95	95
SID	DER	60	13	13.3		95	81

(3) Ethiopia, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
SID	BUR	41	29	29.3		95	91
DER	ALA	49	32	32.0	X	81	91
DER	KEM	49	24	24.0	X	81	99
DER	HAD	42	21	21.3		81	89
DER	SID	60	41	41.3	X	81	95
DER	DER	100	81	100.0		81	81
DER	BUR	37	15	14.7		81	91

(4) Iroquois

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
SEN	SEN	100	83	100.0		83	83
SEN	CAY	72	82	98.8	X	83	80
SEN	ONE	65	30	36.1		83	46
SEN	TUS	50	0	0.0		83	83
CAY	SEN	72	54	67.5		80	83
CAY	CAY	100	80	100.0		80	80
CAY	ONE	73	7	8.7		80	46
ONE	SEN	65	17	37.0		46	83
ONE	CAY	73	18	39.1	X	46	80
ONE	ONE	100	46	100.0		46	46
ONE	TUS	59	0	0.0		46	83
TUS	SEN	50	0	0.0		83	83
TUS	ONE	59	0	0.0		83	46
TUS	TUS	100	83	100.0		83	83

(5) Mazatec

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
HUA	HUA	100	92	100.0		92	92
HUA	JAL	74	35	35.0		92	95
MAT	HUA	94	90	90.0		93	92
MAT	MAT	100	93	100.0		93	93
MAT	JAL	82	33	33.0		93	95
MIG	HUA	94	93	93.0		100	92
MIG	MIG	100	100	100.0		100	100
MIG	JAL	82	56	56.0		100	95
IXC	HUA	78	76	76.0		89	92
IXC	MIG	85	77	77.0		89	100
IXC	IXC	100	89	100.0		89	89
IXC	SOY	85	70	70.0		89	98
IXC	JAL	82	64	64.0		89	95
SOY	HUA	80	73	73.0		98	92
SOY	SOY	100	98	100.0		98	98
SOY	JAL	80	43	43.0		98	95
JAL	HUA	74	73	73.0	X	95	92
JAL	SOY	80	51	51.0	X	95	98
JAL	JAL	100	95	100.0		95	95

(6) Polynesia

TEST	REF	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
HAW	EAS	64	28	34.0		94	96
HAW	HAW	100	94	100.0		94	94
HAW	KAP	49	15	21.0		94	96
HAW	MAN	69	33	39.0		94	98
HAW	MAO	71	25	31.0		94	96
HAW	MAR	70	32	38.0		94	93
HAW	RAR	79	25	31.0		94	93
HAW	SAM	59	25	31.0		94	97
HAW	TAH	76	39	45.0		94	95
HAW	TON	49	3	9.0		94	98
HAW	TUA	77	39	45.0		94	97
HAW	UVE	55	9	15.0		94	96
MAN	EAS	64	48	50.0		98	96
MAN	HAW	69	41	43.0		98	94
MAN	KAP	49	26	28.0		98	96
MAN	MAN	100	98	100.0		98	98
MAN	MAR	73	58	60.0		98	93
MAN	RAR	75	74	76.0		98	93
MAN	SAM	55	24	26.0		98	97
MAN	TON	49	6	8.0		98	98
MAN	TUA	72	96	98.0	X	98	97
MAR	EAS	63	43	50.0		93	96
MAR	HAW	70	50	57.0	X	93	94

(6) Polynesia, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
MAR	KAP	45	59	66.0		93	98
MAR	MAN	73	59	66.0		93	98
MAR	MAR	100	93	100.0		93	93
MAR	RAR	73	58	65.0	X	93	93
MAR	SAM	52	30	37.0	X	93	97
MAR	TON	45	8	15.0		93	98
MAR	TUA	69	97	100.0	X	93	97
RAR	EAS	64	34	41.0		93	96
RAR	HAW	79	30	37.0		93	94
RAR	KAP	54	19	26.0		93	96
RAR	MAN	75	57	64.0	X	93	98
RAR	MAR	73	32	39.0		93	93
RAR	RAR	100	93	100.0		93	93
RAR	SAM	67	19	26.0		93	97
RAR	TON	58	4	11.0		93	98
RAR	TUA	83	90	97.0	X	93	97
SAM	EAS	53	13	16.0		97	96
SAM	HAW	59	26	29.0		97	94
SAM	KAP	53	15	18.0		97	96
SAM	MAN	55	29	32.0		97	98
SAM	MAO	57	28	31.0		97	96
SAM	MAR	52	16	19.0		97	93
SAM	RAR	67	17	20.0		97	93

(6) Polynesia, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
SAM	SAM	100	97	100.0		97	97
SAM	TON	66	16	19.0		97	98
SAM	TUA	62	18	21.0		97	97
SAM	UVE	70	33	36.0		97	96
TAH	EAS	62	30	35.0		95	96
TAH	HAW	76	36	41.0		95	94
TAH	KAP	50	12	17.0		95	96
TAH	MAN	68	39	44.0		95	98
TAH	MAR	67	42	47.0		95	93
TAH	RAR	85	64	69.0		95	93
TAH	SAM	60	22	27.0		95	97
TAH	TAH	100	95	100.0		95	95
TON	EAS	48	15	17.0		98	96
TON	HAW	49	10	12.0		98	94
TON	KAP	45	12	14.0		98	96
TON	MAN	49	24	26.0	X	98	98
TON	MAR	45	10	12.0		98	93
TON	RAR	58	24	26.0	X	98	93
TON	SAM	66	32	34.0	X	98	97
TON	TON	100	98	100.0		98	98
TON	TUA	53	8	10.0		98	97
TON	UVE	86	73	75.0		98	96
TUA	EAS	62	30	33.0		97	96

(6) Polynesia, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
TUA	HAW	77	34	37.0		97	94
TUA	KAP	51	13	16.0		97	96
TUA	MAN	72	56	59.0		97	98
TUA	MAR	69	56	59.0		97	93
TUA	RAR	83	74	77.0		97	93
TUA	SAM	62	23	26.0		97	97
TUA	TON	53	6	9.0		97	98
TUA	TUA	100	97	100.0		97	97

(7) Siouan

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
STO	STO	100	94	100.0		94	94
STO	ASS	89	61	61.0		94	100
STO	MAN	86	23	23.0	X	94	87
STO	NDK	85	46	46.0	X	94	89
STO	LAK	83	10	10.0		94	96
ASS	STO	89	68	68.0		100	94
ASS	ASS	100	100	100.0		100	100
ASS	MAN	94	51	51.0		100	87
ASS	NDK	90	83	83.0		100	89
ASS	LAK	89	50	50.0		100	96
MAN	STO	86	10	10.0		87	94
MAN	ASS	94	84	84.0	X	87	100
MAN	MAN	100	87	100.0		87	87
MAN	NDK	95	82	82.0		87	89
MAN	LAK	91	76	76.0		87	96
NDK	STO	85	22	22.0		89	94
NDK	ASS	90	68	68.0		89	100
NDK	MAN	95	86	86.0		89	87
NDK	NDK	100	89	100.0		89	89
NDK	LAK	90	68	68.0		89	96
LAK	STO	83	3	3.0		96	94
LAK	ASS	89	90	90.0	X	96	100

(7) Siouan, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
LAK	MAN	91	79	79.0		96	87
LAK	NDK	90	90	90.0	X	96	89
LAK	LAK	100	96	100.0		96	96

(8) Trique

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
MIG	SAB	100	99	100.0		99	98
MIG	ITU	84	56	56.0		99	99
MIG	LAG	78	58	58.0		99	98
ITU	SAB	84	92	92.0	X	99	98
ITU	ITU	100	99	100.0		99	99
ITU	LAG	87	98	98.0	X	99	98
LAG	SAB	78	83	83.0	X	98	98
LAG	ITU	87	86	86.0		98	99
LAG	LAG	100	98	100.0		98	98
CHI	SAB	78	74	74.0	X	97	98
CHI	ITU	87	83	83.0		97	99
CHI	LAG	100	97	100.0		97	98
SAB	SAB	100	98	100.0		98	98
SAB	ITU	84	64	64.0		98	99
SAB	LAG	78	57	57.0		98	98

(9) Uganda

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
LUG	LUG	100	79	100.0		79	79
LUG	RUT	64	24	29.8		79	81
LUG	RUN	63	31	39.2		79	82
LUG	LUS	86	49	62.0		79	81
LUG	LUM	54	19	24.1		79	81
RUN	LUG	63	61	74.4	X	82	79
RUN	RUT	86	67	81.7		82	81
RUN	RUN	100	82	100.0		82	82
RUN	RUK	94	72	87.2		82	81
RUN	LUM	49	0	0.0		82	81

(10) Yuman

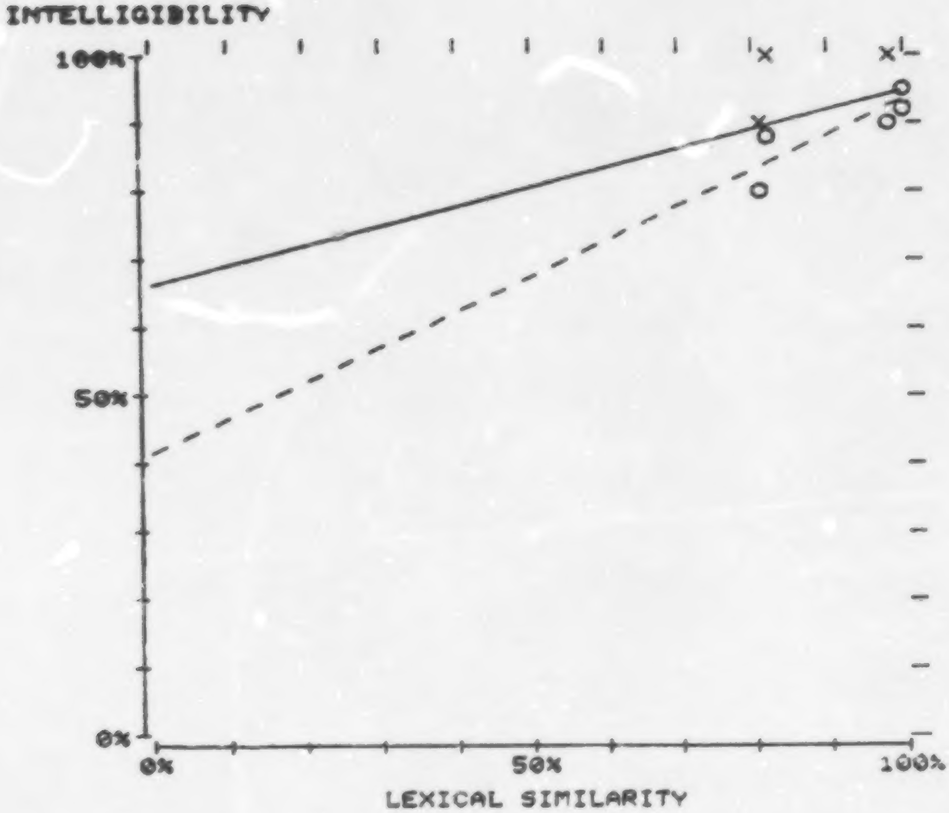
HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
MAR	MAR	100	96	100.0		96	96
MAR	WAL	57	18	22.0		96	96
MAR	YAV	57	13	17.0		96	94
MAR	MOH	85	67	71.0		96	84
MAR	HAV	58	10	14.0		96	91
WAL	MAR	57	14	18.0		96	96
WAL	WAL	100	96	100.0		96	96
WAL	YAV	91	96	100.0	X	96	94
WAL	MOH	63	27	31.0	X	96	84
WAL	HAV	95	91	95.0		96	91
YAV	MAR	57	12	18.0		94	96
YAV	WAL	91	82	88.0		94	96
YAV	YAV	100	94	100.0		94	94
YAV	MOH	62	20	26.0		94	84
YAV	HAV	92	78	84.0		94	91
MOH	MAR	85	77	93.0	X	84	96
MOH	WAL	63	11	27.0		84	96
MOH	YAV	62	13	29.0		84	94
MOH	MOH	100	84	100.0		84	84
MOH	HAV	63	16	32.0		84	91
HAV	MAR	58	11	20.0		91	96
HAV	WAL	95	98	100.0	X	91	96

(10) Yuman, continued

HEAR	SPKR	LEX	INT RAW	INT ADJ	EXCLUDE	SUBJ	TEST
HAV	YAV	92	83	92.0		91	94
HAV	MOH	63	18	27.0		91	84
HAV	HAV	100	91	100.0		91	91

1.3 Scattergrams for raw data

(1) Biliau

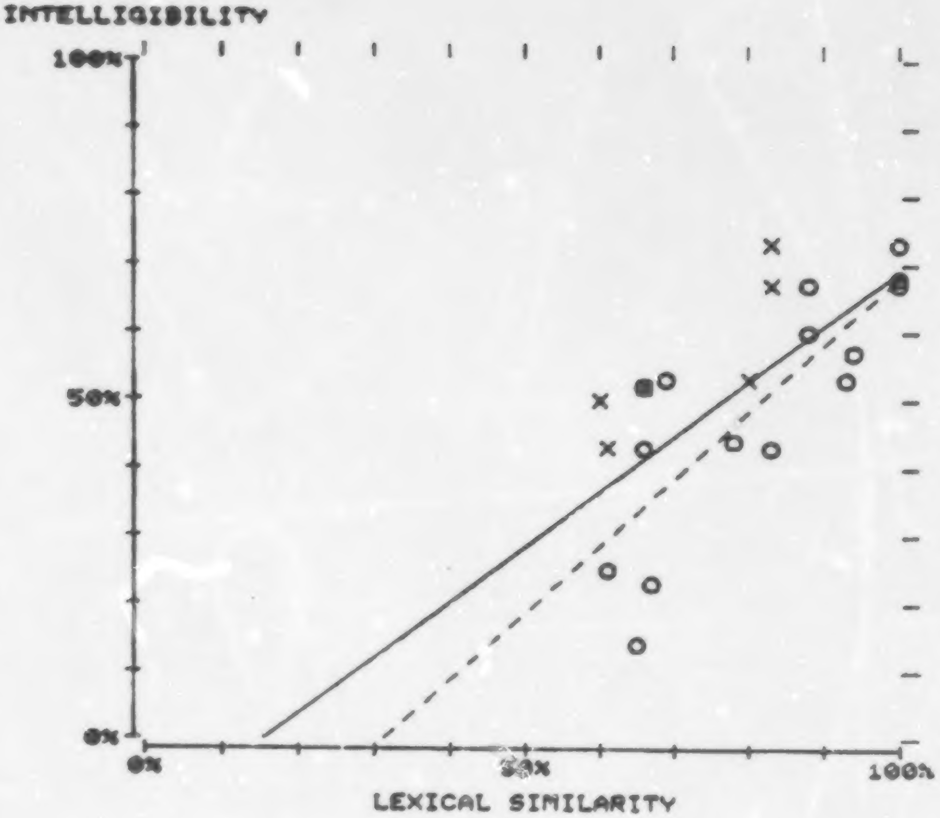


All points (———): Int = .284 Lex + 66.3

Excluding x's (- - -): Int = .519 Lex + 41.5

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	9	18.1	.42487	.2543	6.1	94.7	-233.5
Excl x's	6	74.2	.86156	.0274	3.2	93.4	-79.9

(2) Buang

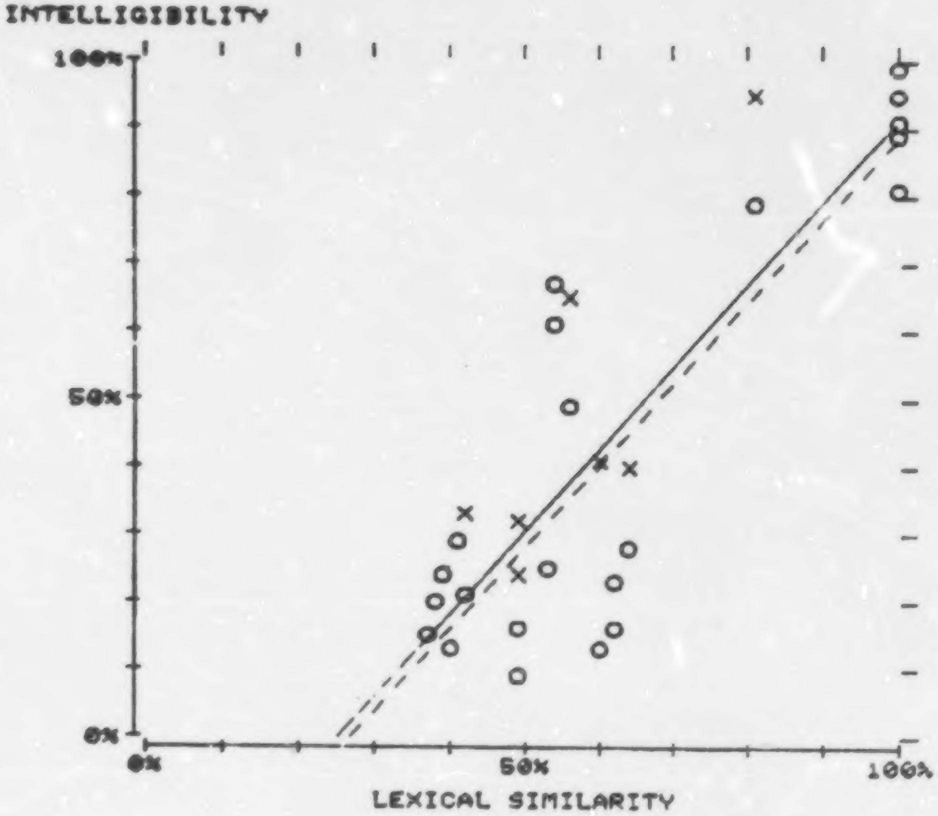


All points (———): Int = .812 Lex - 12.4

Excluding x's (- - -): Int = .988 Lex - 30.8

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	21	49.3	.70232	.0004	11.8	68.8	15.3
Excl x's	15	65.8	.81090	.0002	10.7	68.0	31.1

(3) Ethiopia

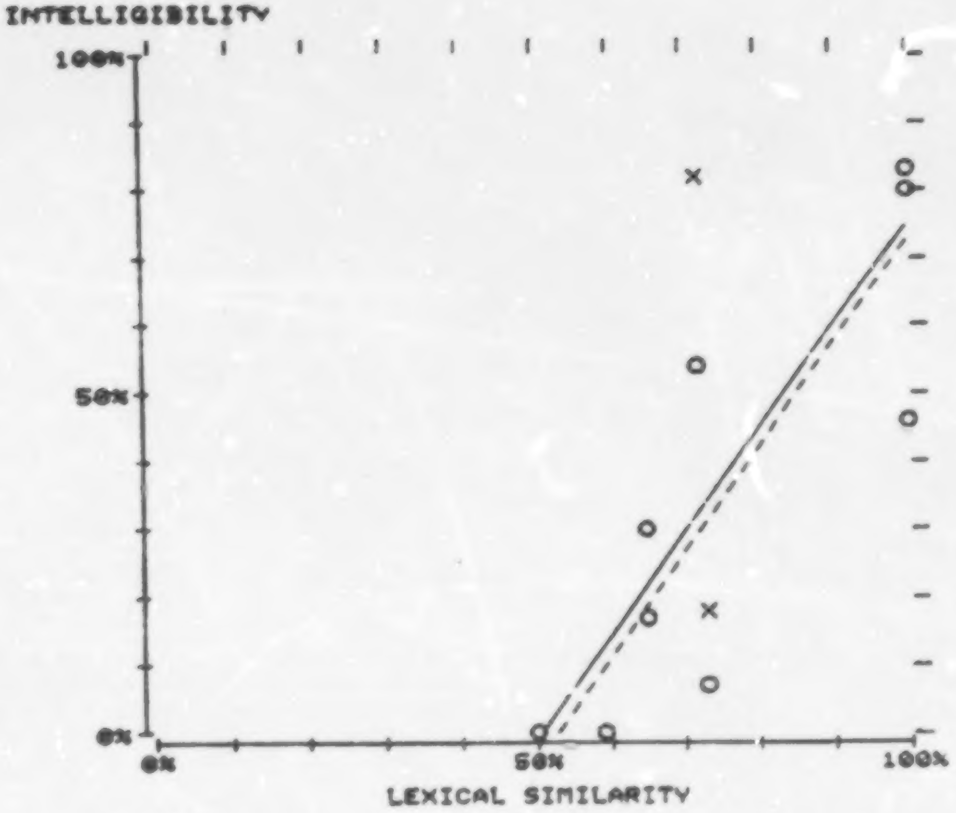


All points (———): Int = 1.217 Lex - 30.5

Excluding x's (- - -): Int = 1.208 Lex - 32.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	30	71.6	.84592	.0001	16.2	91.2	25.1
Excl x's	23	75.1	.86677	.0001	16.1	88.5	26.8

(4) Iroquois

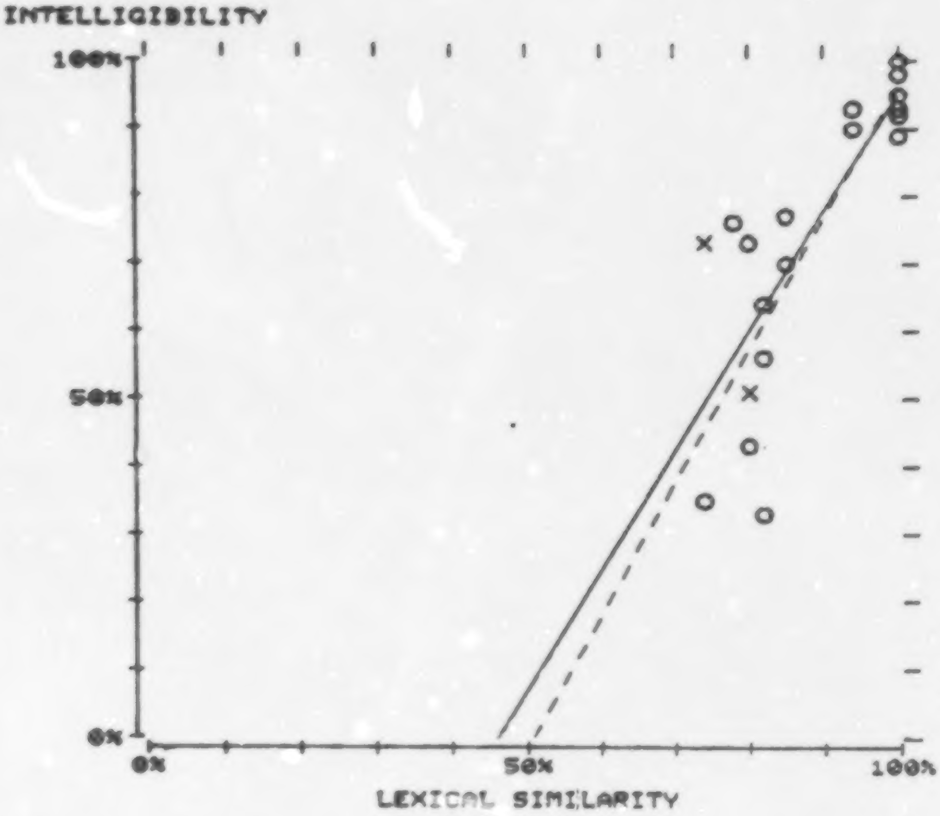


All points (———): Int = 1.519 Lex - 76.9

Excluding x's (- - -): Int = 1.540 Lex - 81.3

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	14	66.0	.81267	.0004	21.0	75.0	50.6
Excl x's	12	80.9	.89944	.0001	15.8	72.7	52.8

(5) Mazatec

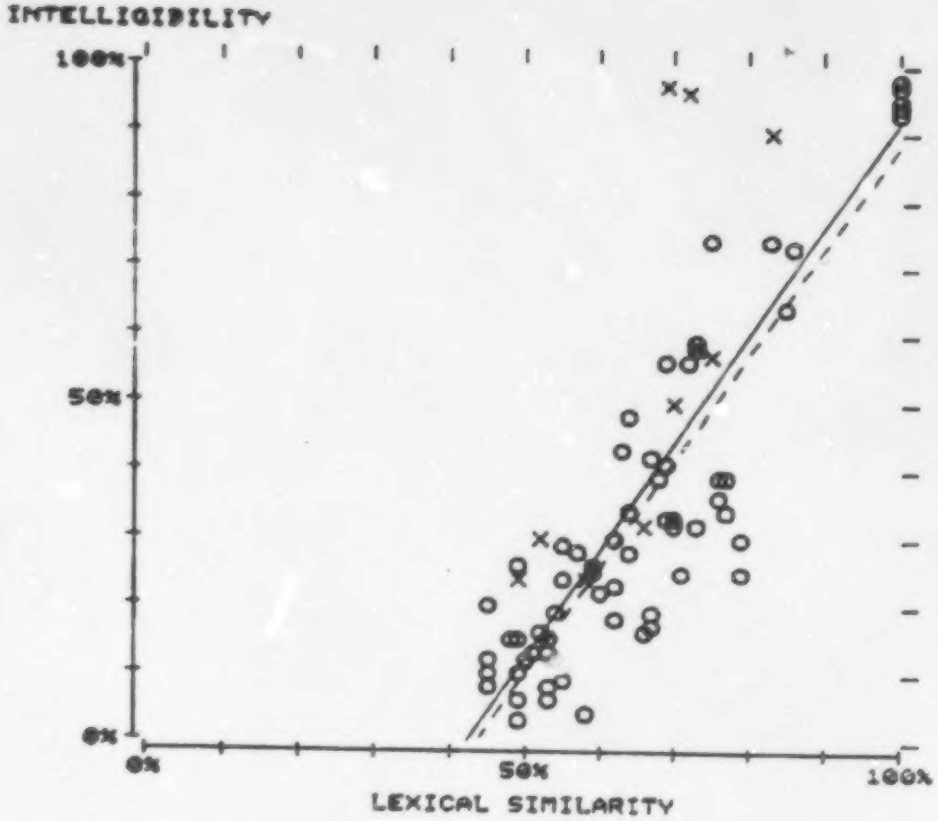


All points (———): Int = 1.766 Lex - 81.5

Excluding x's (- - -): Int = 1.957 Lex - 99.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	19	65.1	.80659	.0001	13.1	95.1	46.1
Excl x's	17	71.7	.84672	.0001	12.1	96.3	50.8

(6) Polynesia

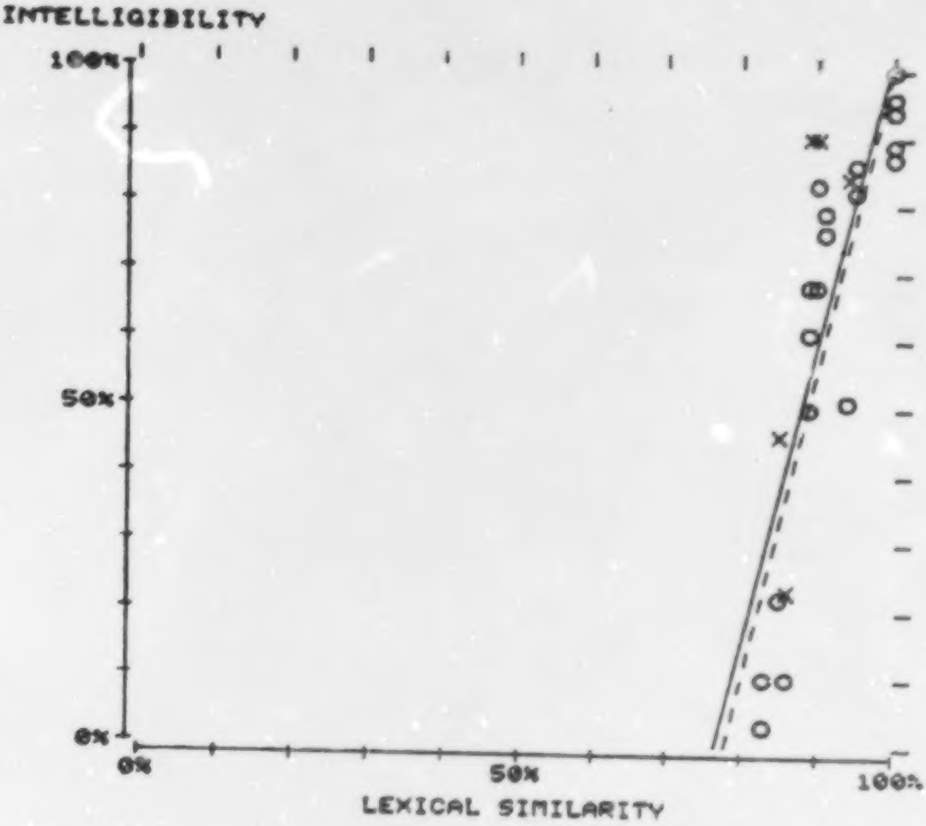


All points (———): Int = 1.588 Lex - 67.2

Excluding x's (- - -): Int = 1.563 Lex - 68.0

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	77	74.6	.86350	.0001	14.4	91.6	42.3
Excl x's	67	83.0	.91091	.0001	11.5	88.3	43.5

(7) Siouan

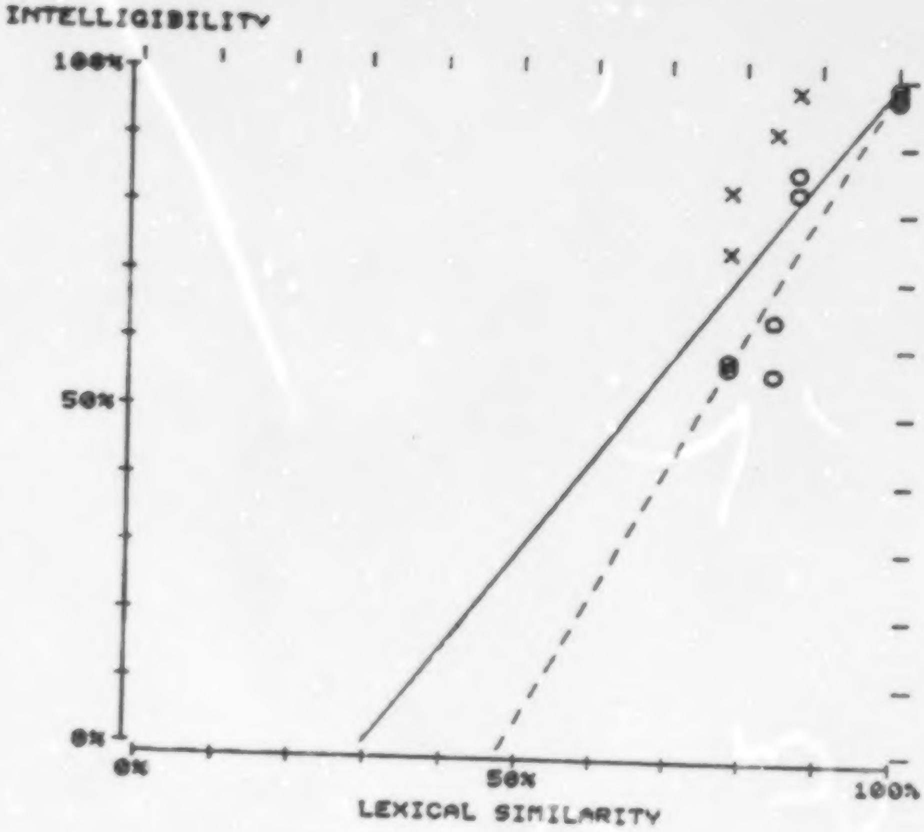


All points (———): Int = 4.385 Lex - 336.0

Excluding x's (- - -): Int = 4.560 Lex - 355.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	25	64.9	.80543	.0001	18.1	102.5	76.6
Excl x's	20	74.2	.86156	.0001	15.9	100.6	77.9

(8) Trique

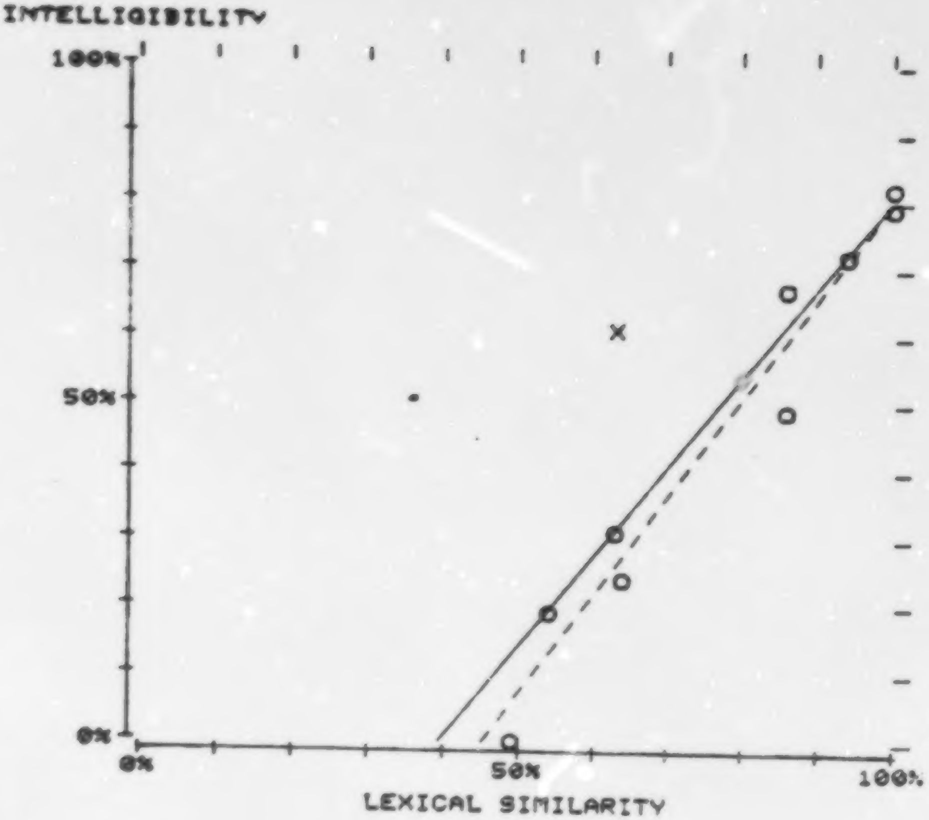


All points (———): Int = 1.405 Lex - 41.3

Excluding x's (- - -): Int = 1.894 Lex - 90.5

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	15	58.5	.76503	.0009	11.2	99.2	29.4
Excl x's	11	88.7	.94174	.0001	6.7	98.9	47.8

(9) Uganda

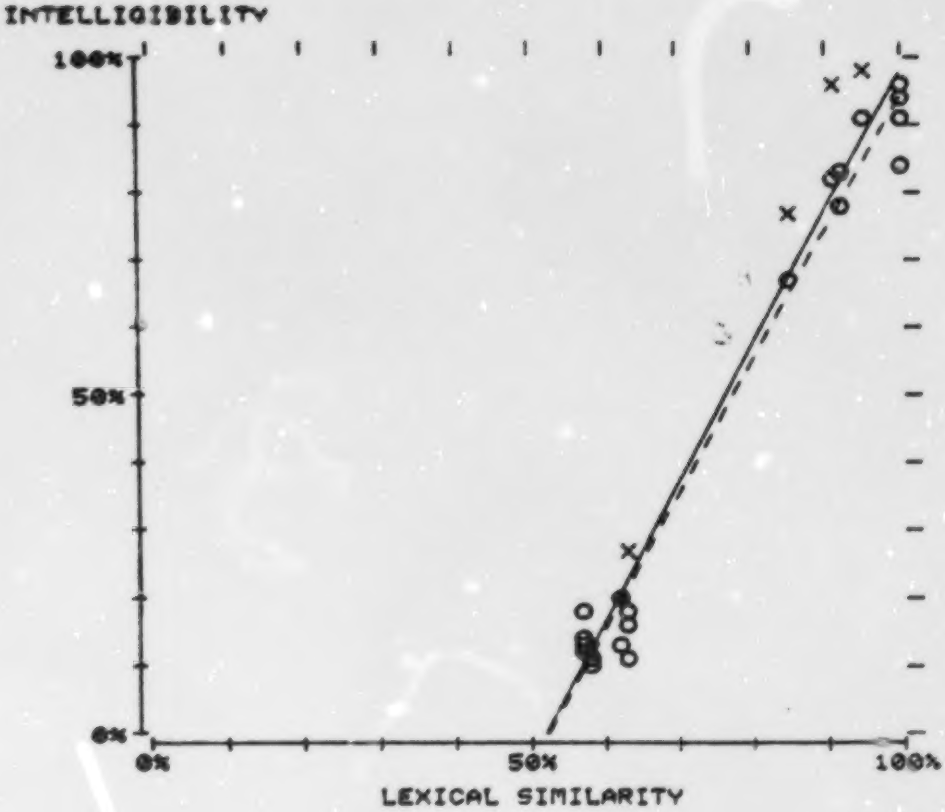


All points (———): Int = 1.325 Lex - 52.2

Excluding x's (- - -): Int = 1.459 Lex - 65.8

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	10	81.8	.90457	.0003	12.8	80.3	39.4
Excl x's	9	96.1	.98010	.0001	6.3	80.1	45.1

(10) Yuman



All points (———): Int = 2.040 Lex - 106.2

Excluding x's (- - -): Int = 1.978 Lex - 103.3

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	25	96.6	.98310	.0001	7.0	97.9	52.0
Excl x's	21	98.1	.99066	.0001	5.2	94.4	52.2

1.4 Adjusting raw intelligibility scores

This section of the appendix sets forth eight tables which were used to select the method of adjusting intelligibility scores for each set of data. Each table has ten rows and six columns. There is one row for each of the ten field studies. The first column is for statistics pertaining to the raw intelligibility scores; the remaining five are for the five different methods of adjusting scores which were tried. See Section 5.2.4 for a description of each adjusting method and the rationale behind each, as well as for the general rationale behind the selection process which is about to be illustrated. In each table, the underlined values indicate the adjustment which was ultimately selected for each data set.

Table 1.1 gives the slope of the regression line for predicting the given measure of intelligibility from lexical similarity. Table 1.2 gives the intercept of the intelligibility axis for these same lines. Thus from these two tables, one can reconstruct the predicting formula for the given set of data and the given type of intelligibility adjustment. That is, predicted percentage of intelligibility equals the slope times the percentage of lexical similarity, plus the intercept. These regression analyses are performed only on the data points which are not suspected of nonsymmetric social factors; only the points plotted as circles in Appendix 1.3 are included.

Table 1.3 reports the percentage of explained variation for each of the regression lines. In Table 1.4 the percentage of explained variation for the raw intelligibility model is subtracted from the percentage for each of the models with adjusted intelligibility. The resulting figures show the net improvement in the ability of the linear model to explain intelligibility after the intelligibility scores are adjusted. A negative value, of course, indicates that the particular adjustment actually lessens the percentage of explained variation. One goal in selecting an adjustment was to find the value in the row which was highest, that is, gave the highest improvement in explained variation.

Table 1.5 reports the percentage of intelligibility which the regression lines predict when similarity is 100%. Table 1.6 shows how much this value deviates from the theoretically expected value of 100%. One goal in selecting an adjustment was to find the value in the row (either positive or negative) which was nearest zero, that is, was nearest the theoretical expectation.

Table 1.7 reports the value of similarity which predicts 0% intelligibility, in other words, the intercepts on the similarity axis. This number gives an idea of how much the predictions from the different model diverge. (We expect all lines to converge on 100%,100% so any differences will appear at the low end of the line.) The mean value of the similarity intercept was computed for all six models concerning the eight field studies which give similar results (Biliau and Siouan are excluded). This yields a mean intercept value of 40.8%. In Table 1.4.8, this mean value is subtracted from all the intercepts in Table 1.4.7. The resulting figures indicate how nearly the given regression line is like all others. One goal in selecting an adjustment was to find the value in the row (either positive or negative) which was nearest zero, that is, which was nearest the overall trend of all studies.

In no case did all three of the stated goals point to the same adjustment. A subjective method of judging the relative importance of the three goals was used to select the adjustments. Objective methods of ranking adjustments and then selecting the adjustment with the highest average rank proved unsatisfactory because they could not account for the qualitative differences between options. For instance, for Buang the hometown adjustment gives the best improvement in explained variation (Table 1.4) and the model with the similarity intercept nearest the average (Table 1.8). On the strength of the highest rank on these two goals, the hometown adjustment turns out to have highest average rank, even though for the deviation from 100% (Table 1.6) it has the second lowest rank. However, a model which deviates by 14.8% from the theoretical expectation of 100% intelligibility for complete similarity is unacceptable. Therefore, the proportional adjustment for subjects which ranks second on the other two goals, but deviates from 100% by only 1.4% was selected.

Table 1.1 Slope

	Raw	Home-town	Subject Prop	Subject Cons	Test Prop	Test Cons
Biliau	0.52	0.79	0.54	0.52	0.59	0.56
Buang	0.99	1.57	1.43	1.02	1.39	0.94
Ethiopia	1.21	1.36	1.31	1.19	1.32	1.17
Iroquois	1.54	2.16	2.10	1.56	2.07	1.55
Mazatec	1.96	2.20	2.00	1.88	2.08	1.98
Polynesia	1.56	1.63	1.63	1.58	1.64	1.59
Siouan	4.56	4.99	4.95	4.59	5.00	4.76
Trique	1.89	1.99	1.93	1.90	1.93	1.90
Uganda	1.46	1.74	1.79	1.44	1.81	1.46
Yuman	1.98	2.11	2.13	1.95	2.15	2.00

Table 1.2 Intelligibility intercept

	Raw	Home-town	Subject Prop	Subject Cons	Test Prop	Test Cons
Biliau	41.5	19.4	45.4	48.3	39.7	42.8
Buang	-30.7	-72.2	-44.5	-2.3	-41.1	4.3
Ethiopia	-32.4	-39.5	-35.1	-23.1	-34.7	-20.3
Iroquois	-81.3	-118.3	-110.2	-55.8	-106.7	-52.6
Mazatec	-99.4	-199.3	-98.6	-86.6	-106.4	-95.8
Polynesia	-68.0	-72.0	-71.2	-64.6	-71.6	-65.4
Siouan	-355.4	-393.4	-386.5	-351.5	-390.8	-366.9
Trique	-90.5	-98.4	-92.2	-89.3	-92.3	-89.3
Uganda	-65.8	-83.0	-80.4	-44.8	-82.1	-47.1
Yuman	-103.3	-111.4	-110.9	-93.8	-112.0	-97.3

Table 1.3 Percentage of explained variation

	Raw	Hometown	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	74.2	77.2	67.9	67.5	78.0	78.0
Buang	65.8	71.2	65.5	64.3	64.2	63.1
Ethiopia	75.1	78.9	75.4	71.5	75.1	74.1
Iroquois	80.9	92.0	88.6	80.4	85.2	81.3
Mazatec	71.7	77.6	67.9	66.6	71.1	70.8
Polynesia	83.0	83.3	83.6	84.3	83.0	83.2
Siouan	74.2	79.7	74.4	74.1	74.5	75.3
Trique	88.7	90.0	87.9	87.7	89.0	88.8
Uganda	96.1	93.5	96.4	96.0	96.2	96.3
Yuman	98.1	98.9	98.9	99.4	98.4	97.8

Table 1.4 Improvement over raw intelligibility

	Raw	Hometown	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	0.	3.0	-6.3	-6.7	3.8	3.7
Buang	0.	5.4	-0.2	-1.4	-1.5	-2.6
Ethiopia	0.	3.8	0.3	-3.6	0.0	-1.1
Iroquois	0.	11.1	7.7	-0.5	4.3	0.4
Mazatec	0.	5.9	-3.8	-5.1	-0.6	-0.9
Polynesia	0.	0.3	0.6	1.4	0.1	0.2
Siouan	0.	5.5	0.2	-0.1	0.3	1.1
Trique	0.	1.3	-0.8	-1.0	0.3	0.1
Uganda	0.	-2.6	0.3	-0.1	0.1	0.2
Yuman	0.	0.7	0.7	1.3	0.2	-0.3

Table 1.5 Predicted intelligibility for 100% similarity

	Raw	Home-town	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	93.4	98.1	99.8	99.8	99.1	99.2
Buang	68.0	85.2	98.6	99.3	98.2	98.6
Ethiopia	88.5	96.0	96.2	95.5	97.1	97.1
Iroquois	72.7	97.6	99.4	100.3	100.1	102.6
Mazatec	96.3	100.9	101.7	101.5	102.0	101.9
Polynesia	88.3	91.0	92.3	93.0	92.7	93.4
Siouan	100.6	105.8	108.6	107.7	109.2	109.0
Trique	98.9	100.6	100.8	100.8	100.5	100.4
Uganda	80.1	90.6	98.8	99.2	99.3	99.4
Yuman	94.4	99.3	101.6	101.6	102.7	102.8

Table 1.6 Deviation from 100% intelligibility

	Raw	Home-town	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	-6.6	-1.9	-0.2	-0.1	-0.9	-0.8
Buang	-32.0	-14.8	-1.4	-0.7	-1.8	-1.4
Ethiopia	-11.5	-4.0	-3.8	-4.5	-2.9	-2.9
Iroquois	-27.3	-2.4	-0.6	0.3	0.1	2.6
Mazatec	-3.7	0.9	1.7	1.5	2.0	1.9
Polynesia	-11.7	-9.0	-7.7	-7.1	-7.3	-6.6
Siouan	0.6	5.8	8.6	7.7	9.2	9.0
Trique	-1.0	0.6	0.8	0.8	0.5	0.4
Ugandan-19.9	-9.4	-1.2	-0.8	-0.7	-0.6	
Yuman	-5.6	-0.7	1.6	1.6	2.7	2.8

Table 1.7 Percentage of similarity for predicted 0% intelligibility

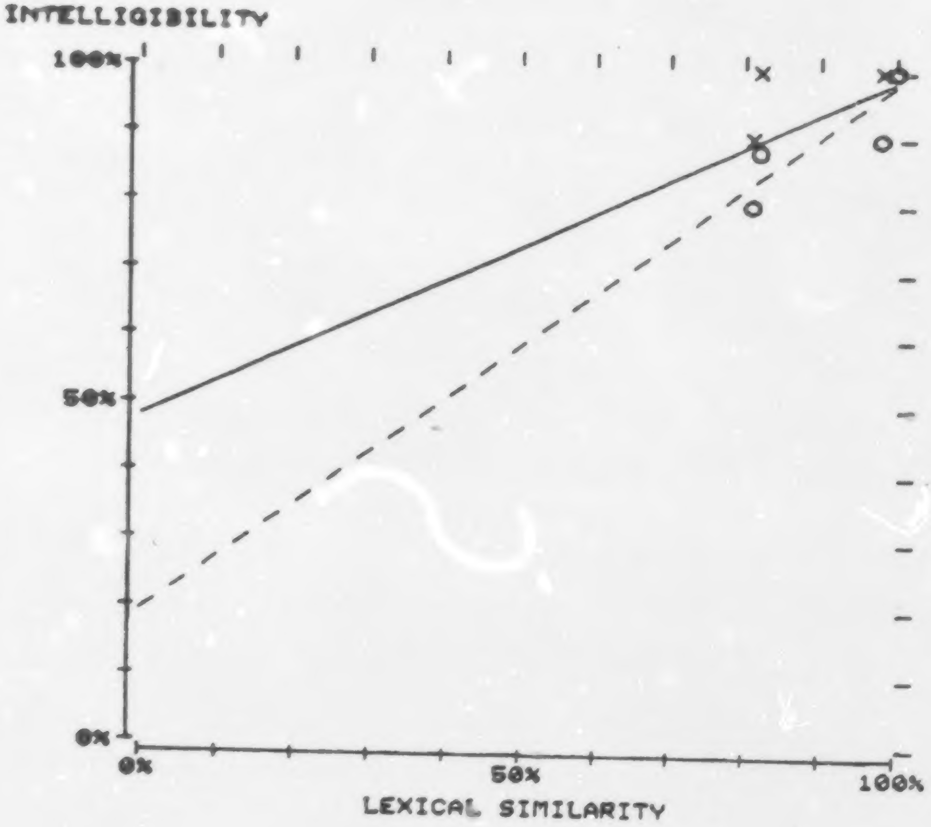
	Raw	Home-town	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	-79.9	-24.6	-83.4	-93.7	-66.8	-75.9
Buang	31.1	45.9	31.1	2.3	29.5	-4.6
Ethiopia	26.8	29.2	26.7	19.5	26.3	17.3
Iroquois	52.8	54.8	49.2	35.8	51.6	33.9
Mazatec	50.8	54.2	49.2	46.0	51.0	48.5
Polynesia	43.5	44.2	43.6	41.0	43.6	41.2
Siouan	77.9	78.8	78.1	76.5	78.2	77.1
Trique	47.8	49.4	47.8	47.0	47.9	47.0
Uganda	45.1	47.8	44.9	31.1	45.3	32.1
Yuman	52.2	52.9	52.2	48.0	52.2	48.6

Table 1.8 Deviation from the mean value of 40.8%

	Raw	Home-town	Subject		Test	
			Prop	Cons	Prop	Cons
Biliau	-120.7	-65.4	-124.2	-134.5	-107.6	-116.7
Buang	-9.7	5.1	-9.7	-38.6	-11.3	-45.4
Ethiopia	-14.0	-11.6	-14.1	-21.3	-14.5	-23.5
Iroquois	12.0	14.0	11.8	-5.0	10.8	-6.9
Mazatec	10.0	13.4	8.4	5.2	10.2	7.7
Polynesia	2.7	3.4	2.8	0.2	2.8	0.4
Siouan	37.1	38.0	37.3	35.7	37.3	36.3
Trique	7.0	8.6	7.0	6.2	7.1	6.3
Uganda	4.3	7.0	4.0	-9.7	4.5	-8.7
Yuman	11.4	12.1	11.4	7.2	11.4	7.8

1.5 Scattergrams for adjusted intelligibility

(1) Biliau

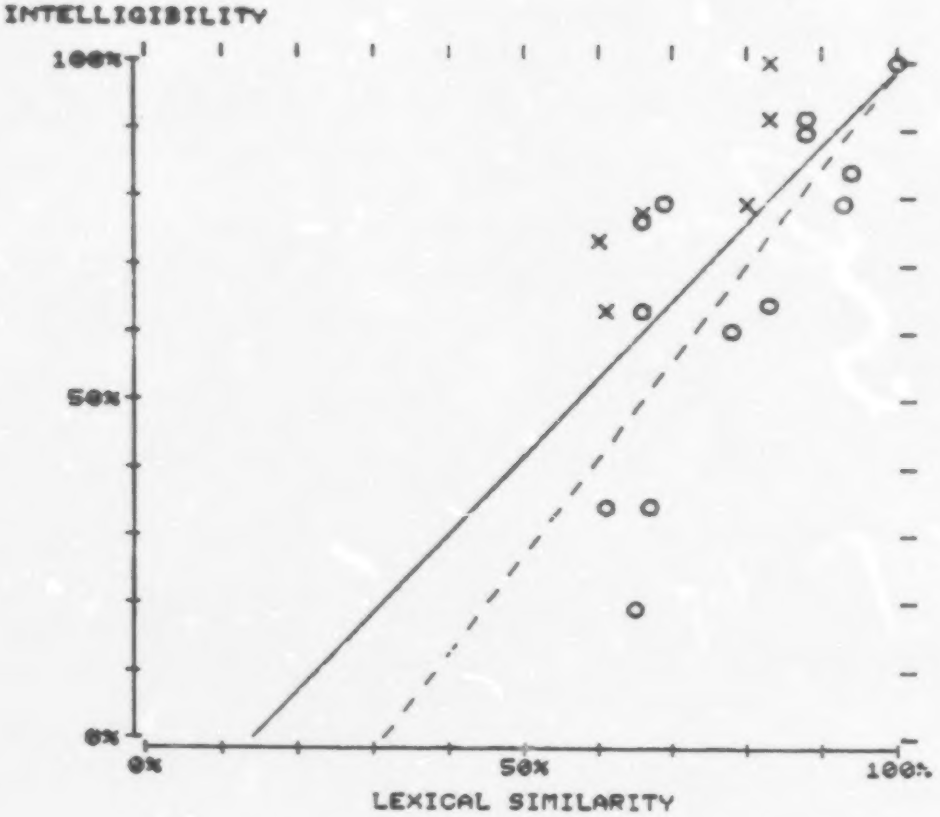


All points (———): Int = .506 Lex + 48.0

Excluding x's (- - -): Int = .788 Lex + 19.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	9	40.6	.63679	.0652	6.1	98.6	-94.8
Excl x's	6	77.2	.87855	.0212	4.5	98.1	-24.6

(2) Buang

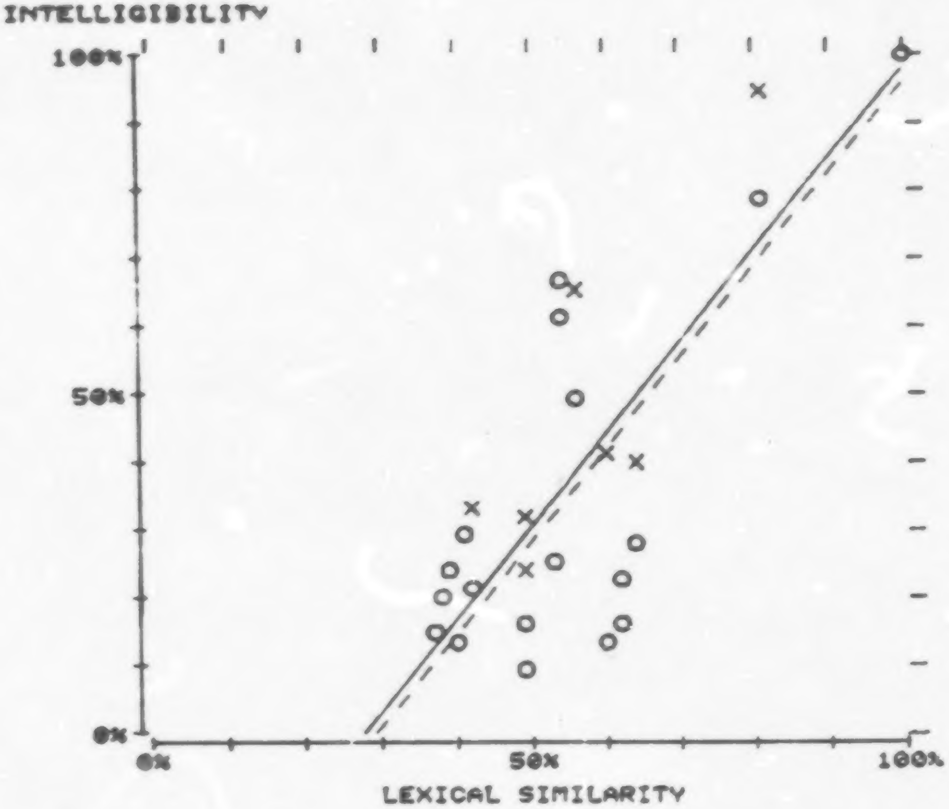


All points (————): Int = 1.148 Lex - 16.0

Excluding x's (- - -): Int = 1.431 Lex - 44.5

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	21	49.3	.70216	.0004	16.7	98.9	13.9
Excl x's	15	65.5	.80949	.0003	15.6	98.6	31.1

(3) Ethiopia

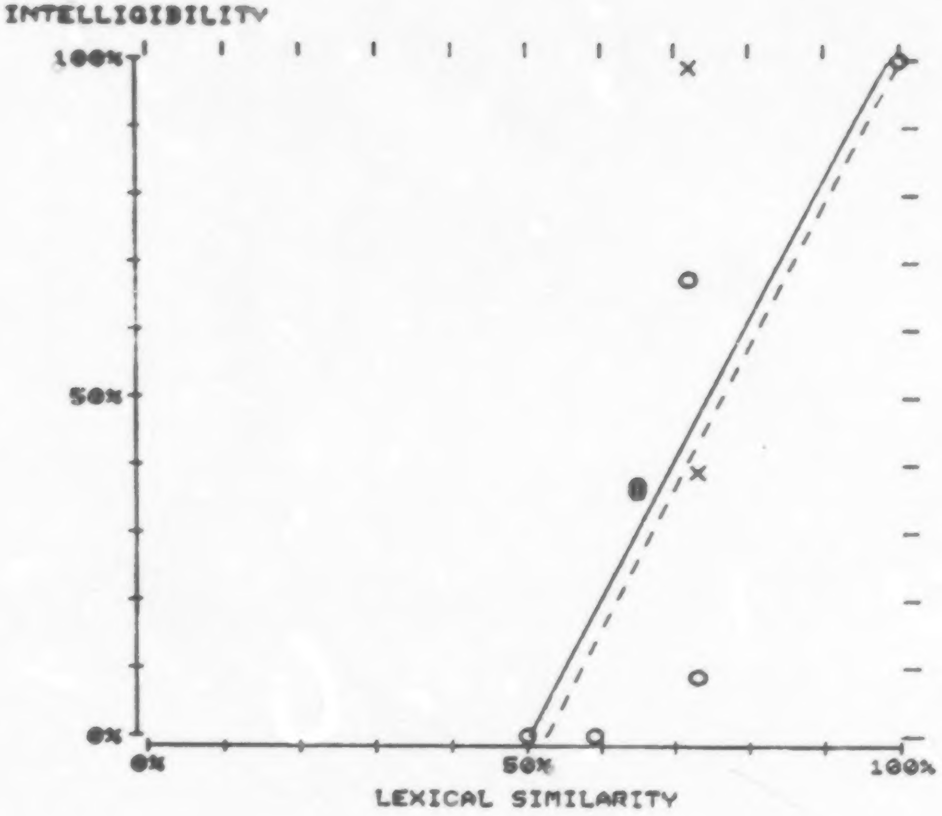


All points (———): Int = 1.354 Lex - 37.3

Excluding x's (- - -): Int = 1.356 Lex - 39.6

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	30	76.0	.87187	.0001	16.0	98.1	27.6
Excl x's	23	78.9	.88827	.0001	16.3	96.0	29.2

(4) Iroquois

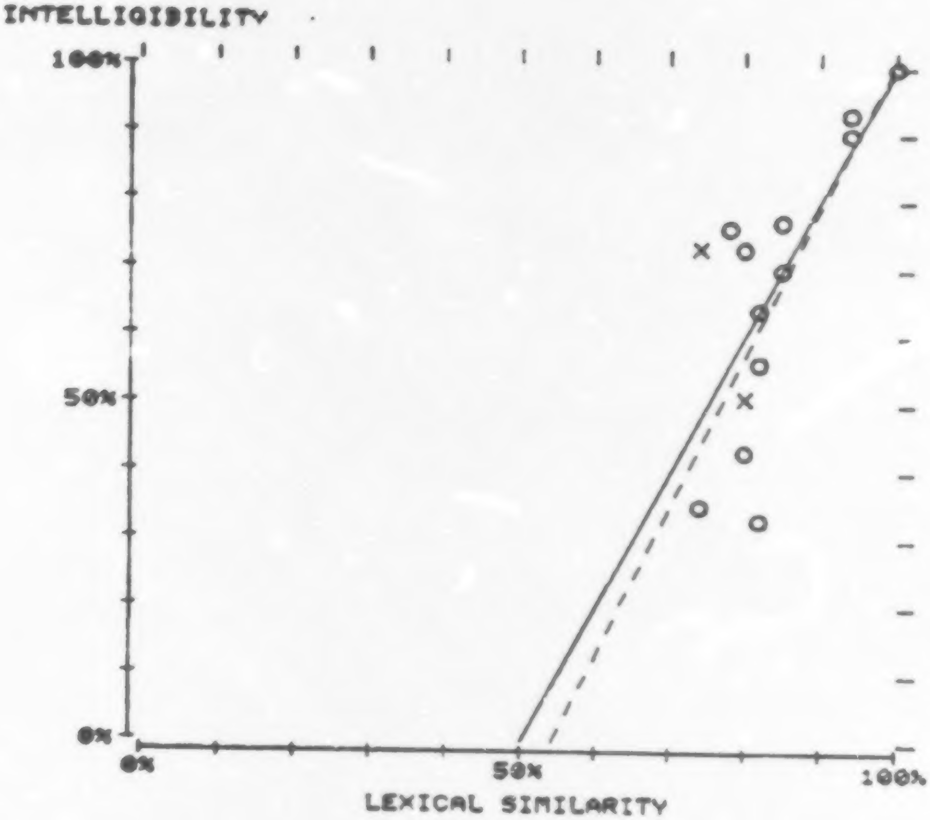


All points (———): Int = 2.069 Lex - 104.3

Excluding x's (- - -): Int = 2.096 Lex - 110.2

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	14	77.1	.87796	.0001	21.8	102.6	50.4
Excl x's	12	88.6	.94131	.0001	15.9	99.4	52.6

(5) Mazatec

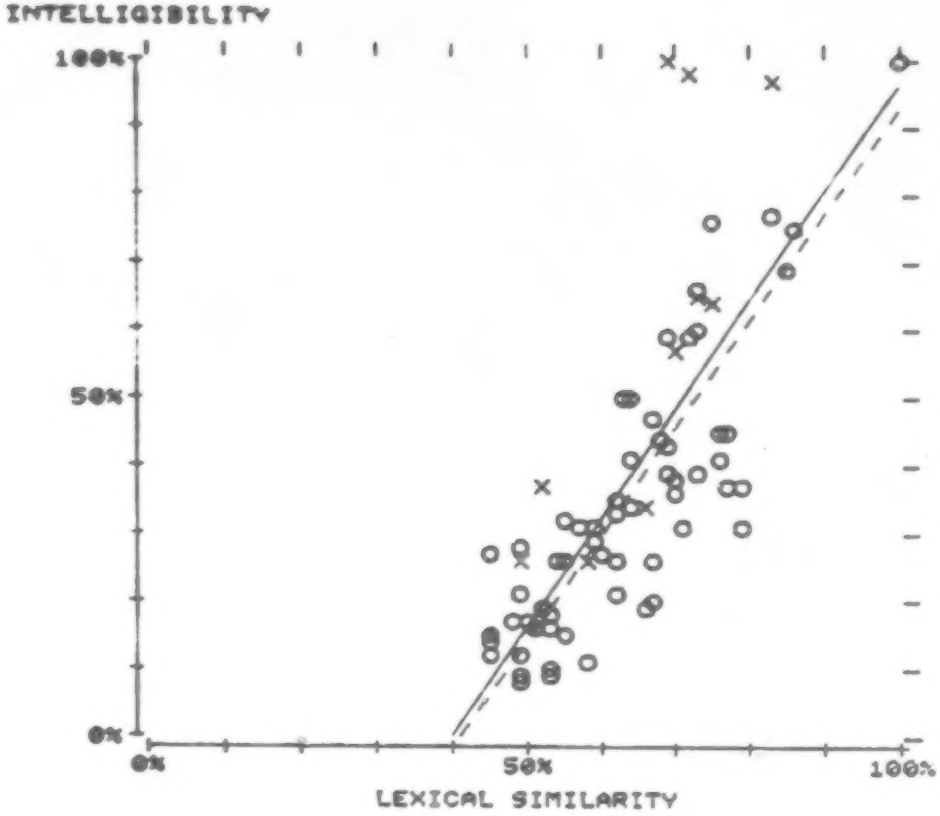


All points (———): Int = 1.995 Lex - 99.8

Excluding x's (- - -): Int = 2.202 Lex - 119.3

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	19	71.0	.84272	.0001	12.9	99.6	50.1
Excl x's	17	77.6	.88111	.0001	11.7	100.9	54.2

(6) Polynesia

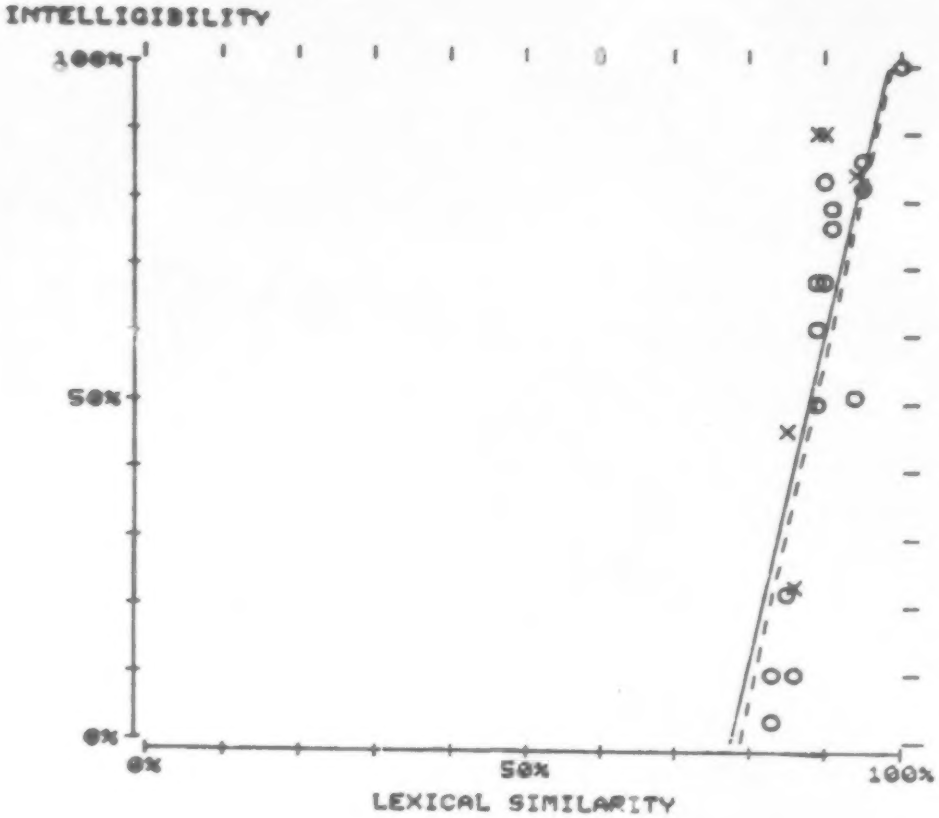


All points (———): Int = 1.605 Lex - 64.0

Excluding x's (- - -): Int = 1.576 Lex - 64.6

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	77	75.8	.87084	.0001	14.1	96.5	39.9
Excl x's	67	84.3	.91840	.0001	11.0	92.9	41.0

(7) Siouan

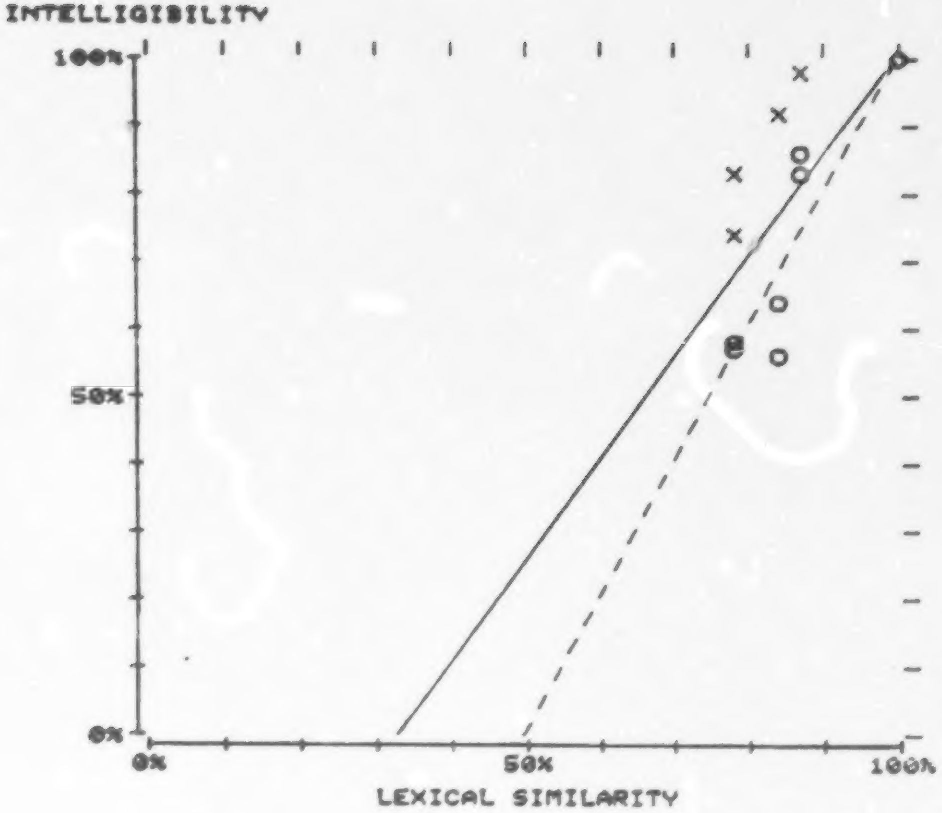


All points (o———): Int = 4.792 Lex - 371.8

Excluding x's (- - -): Int = 4.992 Lex - 393.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	25	70.8	.84143	.0001	17.2	107.4	77.6
Excl x's	20	79.7	.89298	.0001	14.9	105.8	78.8

(8) Trique

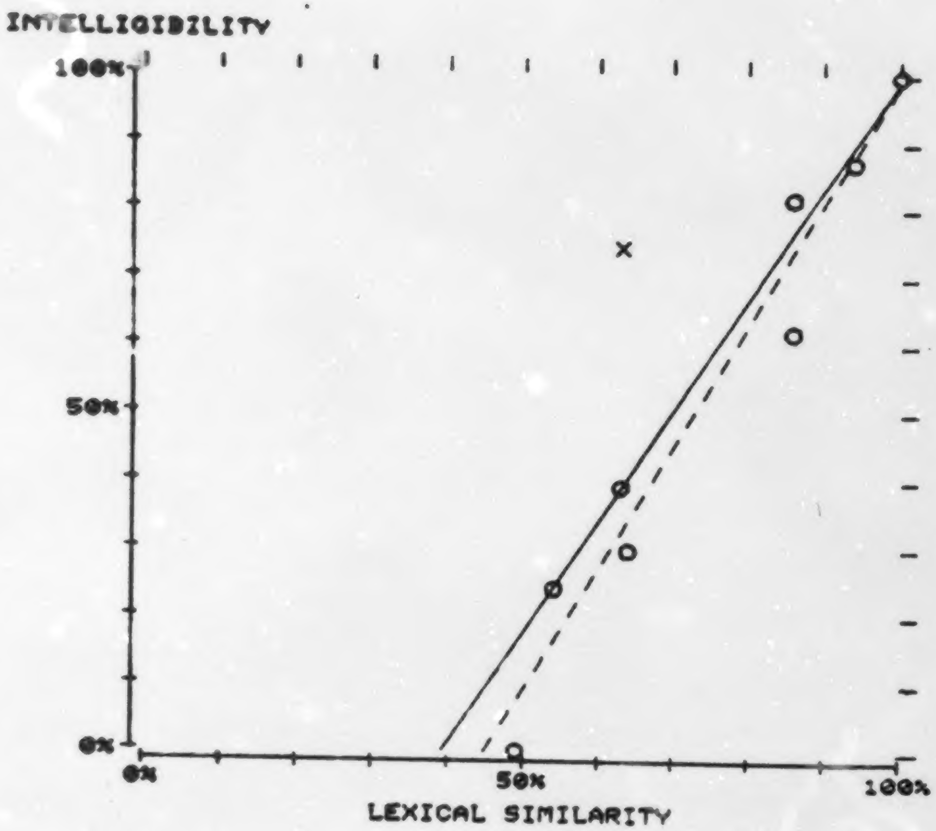


All points (———): Int = 1.495 Lex - 48.6

Excluding x's (- - -): Int = 1.990 Lex - 98.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	15	61.8	.78600	.0005	11.2	100.8	32.5
Excl x's	11	90.0	.94864	.0001	6.5	100.5	49.4

(9) Uganda

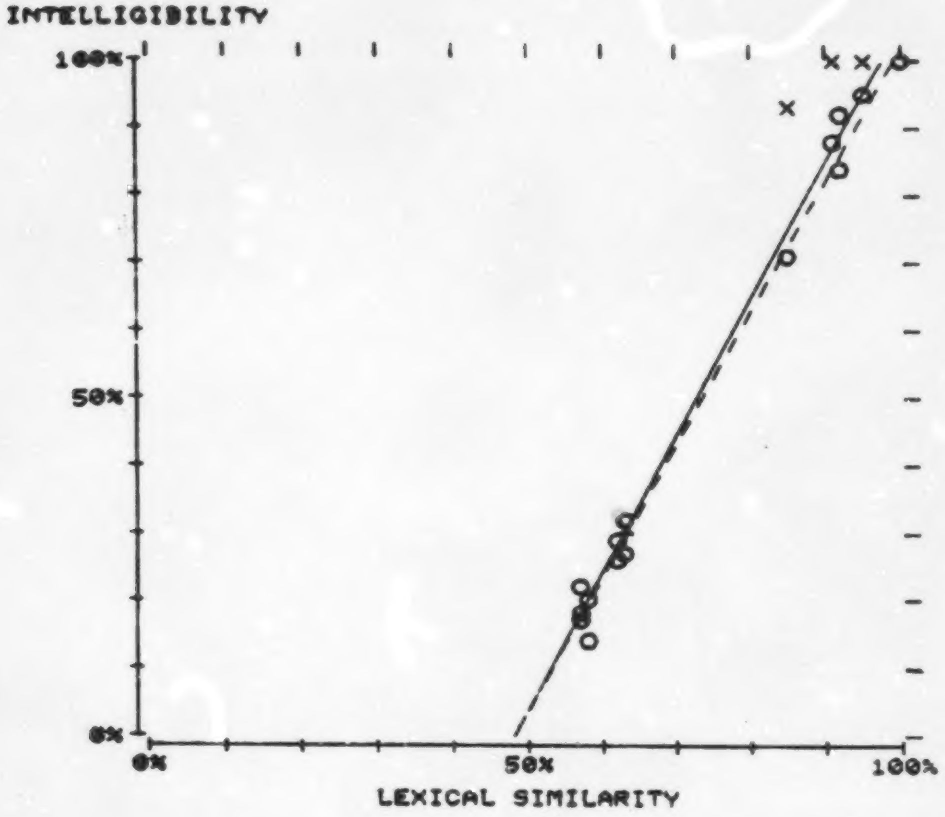


All points (———): Int = 1.632 Lex - 64.0

Excluding x's (- - -): Int = 1.792 Lex - 80.4

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	10	82.7	.90952	.0003	15.3	99.2	39.2
Excl x's	9	96.4	.98174	.0001	7.4	98.8	44.9

(10) Yuman



All points (———): Int = 2.015 Lex - 96.7

Excluding x's (- - -): Int = 1.955 Lex - 93.8

	N	%EV	Corr	Sig	SEE	Lex-100	Int-0
All points	25	97.6	.98772	.0001	5.9	104.8	48.0
Excl x's	21	99.4	.99715	.0001	2.8	101.6	48.0

APPENDIX 2

COMPLETE DATA FOR THE STUDY OF INTELLIGIBILITY ON SANTA CRUZ ISLAND

2.1 Description of the data

The raw data from the Santa Cruz dialect survey are reported in full in Simons 1977a. For the current study those data have been preprocessed in a few ways, primarily in order to give uniform dimensions and dialect labels to all data tables. In the raw data, different sets of data have different numbers of rows and columns or different row and column labels. For the present study, many of those rows and columns are combined and many are renamed to make all data tables uniform and comparable.

Thirteen dialects are used in both dimensions of all tables. These are the thirteen points at which intelligibility was tested. The first section of this appendix gives the mnemonic codes for the thirteen test points and a listing of the villages they represent. In the sections that follow all the data used in the analysis are described and listed.

2.1.1 The thirteen dialects

The thirteen dialects used throughout this study are listed below. Each is viewed as a unique dialect made up of one or more villages. When a number of villages are combined, the villages are near neighbors and their speech varieties are identical or very nearly so. The term dialect is used loosely here. It makes no suggestion of how different the speech varieties are; it only implies that the speech communities are in some way distinct, either spatially or linguistically, or both.

Figures 6.1 and 6.2 in Section 6.1.3 give sketch maps of the island showing the location of the dialects. It should be noted that the villages along the northeast shore of the island are omitted. This is because they are small and are all recent migrations from more populous villages which are included in the study. Likewise, the eastern tip of the island, which does not appear in the maps, is inhabited only by recent immigrants from another island. The mnemonic codes for the dialects and the villages they represent are as follows:

(1) NEO = Neo
(2) MAT = Matu
(3) BAN = Mbanua, Noole, Lwepe, Moneu, Monao, Nou, Uta
(4) NEP = Nepa, Palo, Mbalo, Mateone, Nepu, Io, Napo
(5) LWO = Lwowa, Malo
(6) VEN = Venga
(7) NEM = Nemba
(8) BYO = Mbanyo, Manoputi, Manamini
(9) NOP = Noepe, Mbapo, Monan
(10) NEA = Nea, Nemboi

195

(11) NOO = Nooli, Nonia, Mbonembwe
(12) MBI = Mbimba
(13) NNG = Nanggu, Utongo

2.1.2 Population

The population of the thirteen dialects is as follows:

NEO	200
MAT	120
BAN	450
NEP	320
LWO	370
VEN	290
NEM	180
BYO	180
NOP	140
NEA	220
NOO	280
MBI	140
NNG	200
Total	3090

2.1.3 Geographic distance

The distance between dialects is measured between their main villages. The main village is the one listed first in the list just given. Distance is measured in terms of the number of minutes required to travel between the dialects. These figures must be viewed as approximations at best. In most cases, they are walking times. In the case of Neo, Matu, Mbimba, and Nanggu, boating (either sailing or paddling or both) is involved for certain stretches. With the recent advent of roads, vehicles, and outboard motors within the past one or two decades, many of these distances have been shortened. However, such means of transportation are still not available to everyone.

Table 2.1 shows the minutes of traveling time between the dialects. The figures on the diagonal, which represent the distance from a dialect to itself, are an approximation to the radius of the dialect. When the dialect consists of only a single village, the distance is given as 5 minutes. If it consists of two or more villages, the average distance from the central village to the others is given.

The last column of Table 2.1 gives the average distance from a speaker of the given dialect to all other inhabitants of the island. A simple average of the distance to all dialects could have been computed by summing the figures in a row and dividing by thirteen. However, such a statistic does not take into account the differing populations of the dialects. Therefore the average distance from an individual to all other individuals is computed. This is done by multiplying each distance in a row by the population of the dialect for the column. Then the row is summed and divided by the total population of the island, that is, 3090. The result is the average distance separating an individual from that dialect from all other individuals on the island. For instance, for BAN the average distance is 180 minutes. One way to interpret this figure is that by traveling no more than 180 minutes from home, a BAN person could come into contact with half of the residents of the island.

Table 2.1 Geographic distance

Values are minutes of traveling time

	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG	Mean
NEO	5	120	155	185	95	155	260	330	435	345	415	775	805	276
MAT	120	5	180	210	150	210	315	385	490	370	440	800	830	313
BAN	155	180	10	30	60	70	175	245	295	190	260	620	650	180
NEP	185	210	30	15	90	100	205	275	265	160	230	590	620	185
LWO	95	150	60	90	15	60	165	235	340	250	320	680	710	200
VEN	155	210	70	100	60	5	105	175	280	260	330	690	720	202
NEM	260	315	175	205	165	105	5	70	175	280	415	775	805	262
BYO	330	385	245	275	235	175	70	30	105	210	345	705	735	281
NOP	435	490	295	265	340	280	175	105	10	105	240	600	630	299
NEA	345	370	190	160	250	260	280	210	105	10	135	495	525	238
NOO	415	440	260	320	320	330	415	345	240	135	15	360	390	292
MBI	775	800	620	590	680	690	775	705	600	495	360	5	30	562
NNG	805	830	650	620	710	720	805	735	630	525	390	30	5	588

island.

In Table 2.2, the distances in each row of Table 2.1 are divided by the average distance for that dialect and then multiplied by 100 to convert it to a percentage. The result is a table of distances measured relative to the perspective of each dialect (see Section 6.1.3). The rows are labeled "From:" while the columns are labeled "To:". For instance, the distance from Nanggu (NNG) to Mbanua (BAN) is only 110% of the average distance from NNG, while from BAN to NNG it is 360% of the average distance from BAN. This suggests that from an insider's perspective, a NNG speaker views BAN as being nearer to his own village than a BAN speaker would view the distance to NNG. Note that the distances in Table 2.1 are symmetric (that is, the same in both directions), while those in Table 2.2 are not.

Table 2.2 Relative geographic distance

Values are percentage of mean distance from the origin point

		To:												
		NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
	NEO	2	43	56	67	34	56	94	120	158	125	150	281	292
	MAT	38	2	58	67	48	67	101	123	157	118	141	256	265
	BAN	86	100	6	17	33	39	97	136	164	105	144	344	360
	NEP	100	113	16	8	49	54	111	148	143	86	173	318	335
	LWO	48	75	30	45	8	30	83	118	170	125	160	340	355
	VEN	77	104	35	50	30	2	52	87	139	129	164	342	357
From:	NEM	99	120	67	78	63	40	2	27	67	107	158	295	307
	BYO	117	137	87	98	84	62	25	11	37	75	123	251	261
	NOP	146	164	99	89	114	94	59	35	3	35	80	201	211
	NEA	145	156	80	67	105	109	118	88	44	4	57	208	221
	NOO	142	151	89	79	100	113	142	118	82	46	5	123	134
	MBI	138	142	110	105	121	123	138	125	107	88	64	1	5
	NNG	137	141	110	105	121	122	137	125	107	89	66	5	1

2.1.4 Density of population

n Appendix 2.1.2, population was measured in absolute terms. It can also be measured relatively with respect to the whole dialect system (Section 6.1.3) by computing density of population. When population is viewed in terms of its density, rather than in absolute numbers, one is hypothesizing that the attraction of a dialect could be enhanced by the nearness of its neighbors; likewise, motivation for its speakers to travel widely to engage in contact might be diminished.

Here density is computed roughly in terms of people per square mile. Actually, no miles are measured. Rather, the traveling distances in minutes are divided by twenty to give a rough approximation to miles. The density at a dialect is computed as the density in the square mile in which the dialect is located. The contribution of the dialect itself is arbitrarily set at its population (even when it may cover more than a square mile). The contribution of the other dialects is computed as follows. The population of another dialect is viewed as evenly distributed over a circular area which has a radius equal to the distance between the two dialects. This distance is squared and then multiplied by pi (3.1416) to compute the density at the first dialect. For each dialect the contributions of the other twelve are computed and added to the population of the original community. The result is a measure of population density at each dialect. The results are as follows:

NEO	212
MAT	128
BAN	520
NEP	397
LWO	407
VEN	324
NEM	194
BYO	191
NOP	148
NEA	229
NOO	285
MBI	169
NNG	221

2.1.5 Lexical similarity

Lexical similarity between the dialects was measured as a gauge of their linguistic similarity. The computation of cognate percentages is based on the Swadesh 100-word list. The lists were collected by Richard Buchan and are reproduced in full in Simons 1977a. The percentage of lexical cognates between all the speech communities is given in Table 2.3. Items were judged cognate simply on the basis of phonetic similarity. No attempt was made to distinguish between direct inheritance and indirect inheritance through borrowing.

2.1.6 Lexical distance

The linguistic distance between dialects is approximated by computing lexical distance. Lexical distance is the percentages of basic vocabulary that is not cognate. This is computed by subtracting the cognate percentages in Table 2.3 from 100%.

The lexical distance between dialects is given in Table 2.4. In the last column of the table, the average lexical distance separating an individual of each dialect from all other individuals on the island is given. This average distance is computed just as described for geographic distance in Appendix 2.1.3. In Table 2.5 the lexical distance figures in each row are divided by the average distance for the row to derive a relative, nonsymmetric measure of linguistic distance. The interpretation of these figures is analogous to the interpretation discussed in Appendix 2.1.3 for relative geographic distance.

2.1.7 Measured intelligibility

Intelligibility between dialects was measured using the technique described in Section 2.1. The responses were scored on the four point scale described in Section 2.1.4. The results of the intelligibility testing are displayed in Table 2.6. The responses given are what I judged to be the norms for the dialects taking the test. When an individual having close contact with the dialect on the test tape dominated the beginning of a test, I directed questions to other members of the group in order

Table 2.3 Lexical similarity

Values are percentage of cognates

	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
NEO	100	87	85	83	87	86	78	70	68	65	59	59	50
MAT	87	100	95	86	97	95	85	75	72	68	63	63	53
BAN	85	95	100	87	96	93	85	77	74	72	65	65	54
NEP	83	86	87	100	89	87	83	75	78	74	66	66	54
LWO	87	97	96	89	100	98	87	77	74	72	65	65	54
VEN	86	95	93	87	98	100	86	76	72	70	63	63	53
NEM	78	85	85	83	87	86	100	84	78	75	70	70	59
BYO	70	75	77	75	77	76	84	100	88	80	73	73	63
NOP	68	72	74	78	74	72	78	88	100	88	78	78	64
NEA	65	68	72	74	72	70	75	80	88	100	85	85	68
NOO	59	63	65	66	65	63	70	73	78	85	100	100	72
MBI	59	63	65	66	65	63	70	73	78	85	100	100	72
NNG	50	53	54	54	54	53	59	63	64	68	72	72	100

Table 2.4 Lexical distance

Values are percentage of non-cognates

	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG	Mean
NEO	0	13	15	17	13	14	22	30	32	35	41	41	50	23.1
MAT	13	0	5	14	3	5	15	25	28	32	37	37	47	17.8
BAN	15	5	0	13	4	7	15	23	26	28	35	35	46	16.8
NEP	17	14	13	0	11	13	17	25	22	26	34	34	46	19.0
LWO	13	3	4	11	0	2	13	23	26	28	35	35	46	15.9
VEN	14	5	7	13	2	0	14	24	28	30	37	37	47	17.4
NEM	22	15	15	17	13	14	0	16	22	25	30	30	41	19.3
BYO	30	25	23	25	23	24	16	0	12	20	27	27	37	22.8
NOP	32	28	26	22	26	28	22	12	0	12	22	22	36	23.1
NEA	35	32	28	26	28	30	25	20	12	0	15	15	32	23.7
NOO	41	37	35	34	35	37	30	27	22	15	0	0	28	27.6
MBI	41	37	35	34	35	37	30	27	22	15	0	0	28	27.6
NNG	50	47	46	46	46	47	41	37	36	32	28	28	0	38.7

Table 2.5 Relative lexical distance

Values are percentage of mean distance from origin point

To:

From:		NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
	NEO	0	56	65	74	56	61	95	130	139	152	177	177	216
	MAT	73	0	28	79	17	28	84	140	157	180	208	208	264
	BAN	89	30	0	78	24	42	89	137	155	167	209	209	274
	NEP	90	74	69	0	58	69	90	132	116	137	179	179	243
	LWO	82	19	25	69	0	13	82	145	164	176	221	221	290
	VEN	80	29	40	75	11	0	80	138	161	172	213	213	270
From:	NEM	114	78	78	88	67	73	0	83	114	130	156	156	213
	BYO	131	110	101	110	101	105	70	0	53	88	118	118	162
	NOP	138	121	112	95	112	121	95	52	0	52	95	95	156
	NEA	148	135	118	110	118	126	105	84	51	0	63	63	135
	NOO	149	134	127	123	127	134	109	98	80	54	0	0	102
	MBI	149	134	127	123	127	134	109	98	80	54	0	0	102
	NNG	129	121	119	119	119	121	106	96	93	83	72	72	0

to assess how well the majority was understanding. This latter assessment is reported in the table of scores. The periods indicate that intelligibility was not tested for that particular pairing of dialects. The dialects listed along the left hand side of the table are those which listened to the test tapes. Those listed along the top are those which were the speakers on the test tapes. Thus, the "2" in the top row of the table means that the people from Neo scored partial intelligibility when they listened to the dialect of Nea.

2.1.8 Opinions about intelligibility

Before intelligibility tests were given, the members of the group were asked how well they understood the other dialects on Santa Cruz. The question asked was: "How much of the speech of village X do you understand?" The answers were scored on a three point scale: 2 = all of it, 1 = some of it, 0 = none of it. The

Table 2.6 Measured intelligibility

3 = Full intelligibility
2 = Partial intelligibility
1 = Sporadic recognition

Dialect of speaker:

Dialect of hearers:	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
NEO	3	.	.	3	.	.	.	.	.	2	.	.	.
MAT	3	.	3	.	.	.	3	2	.	.	1	.	1
BAN	3	.	3	3	.	.	.	2	2	2	2	.	1
NEP	3	.	3	3	.	.	3	3	3	3	2	.	1
LWO	3	.	.	.	3	.	.	3	3	2	2	.	.
VEN	3	.	.	.	.	3	3	3	3	2	2	.	1
NEM	.	.	.	.	.	.	3	.	.	3	2	.	1
BYO	3	.	.	.	.	.	.	3	.	3	3	.	1
NOP	3	.	.	3	.	.	.	3	3	3	3	.	1
NEA	2	.	.	3	.	.	.	.	.	3	3	.	1
NOO	2	.	3	.	.	.	.	3	.	3	3	.	3
MBI	2	.	.	2	.	.	.	3	.	3	3	.	3
NNG	3	.	3	.	.	.	.	3	.	.	3	.	3

results of this investigation are given in Table 2.7. The dialects listed along the left hand side of the table are those to which the question was asked. Those listed along the top are the ones which were asked about. Thus the score of "0" in the top row of the table indicates that the people of Neo said they could not understand any of the speech of Nanggu.

The bottom row and the rightmost column of the table give the attraction and motivation of the dialects as indicated by these opinions (see Section 6.1.2.3). These are weighted by population in the same manner as the average geographic and lexical distance. That is, they are computed as an average intelligibility per individual, rather than per dialect. This is done by multiplying the opinion scores by the population of the intersecting dialect, summing, and dividing by total

Table 2.7 Opinions about Intelligibility

"How much of the speech of village X do you understand?"

2 = understand all of it
1 = understand some of it
0 = understand none of it

Dialect asked about:

Dialect asked:

	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG	Mot
NEO	2	2	2	2	2	2	2	2	2	2	2	2	0	.93
MAT	2	2	2	2	2	2	1	1	0	0	0	0	0	.61
BAN	2	2	2	2	2	2	2	1	1	1	1	1	0	.74
NEP	2	2	2	2	2	2	2	2	2	2	2	2	1	.96
LWO	2	2	2	2	2	2	2	2	2	2	2	2	0	.93
VEN	2	2	2	2	2	2	2	2	2	2	2	2	1	.96
NEM	2	2	2	2	2	2	2	2	2	2	2	2	1	.97
BYO	2	2	2	2	2	2	2	2	2	2	2	2	1	.97
NOP	2	2	2	2	2	2	2	2	2	2	2	2	1	.97
NEA	2	2	2	2	2	2	2	2	2	2	2	2	1	.97
NOO	2	2	2	2	2	2	2	2	2	2	2	2	1	.96
MBI	1	2	2	2	2	2	2	2	2	2	2	2	2	.97
NNG	2	2	2	2	2	2	2	2	2	2	2	2	2	1.0
Att	.98	1.0	1.0	1.0	1.0	1.0	.98	.90	.88	.88	.88	.88	.33	

population. It should be noted that these computations do not include the diagonal in the martix; they refer only to the other twelve dialects. The scores are further divided by two in order to convert them to a range of zero to one and make them easier to interpret. The attraction figures (Att) can be interpreted as the proportion of the island's population which claim to understand the given dialect. Thus we see that 100% of the islanders claim to understand Mbanua (BAN) while only 33% claim to understand Nanggu (NNG). The motivation figures (Mot) can be interpreted as the proportion of the island's population which the given dialect claims to understand. Thus, we see that NNG claims to understand 100% of the islanders while BAN claims to understand only 74%.

2.1.9 Contact through church festivals

With the exception of some small newly established settlements and Graciosa Bay, where four churches serve the 14 villages, every village on Santa Cruz has a church. Each church takes its name from a saint or a feast day within the church year (e.g. Resurrection, Trinity, Ascension). Once a year, on the appointed day of its saint or feast, each church holds a festival. The festival begins with a special communion service in the church. This is followed by feasting and dancing which continues all night. The young people participate in sports competitions as well. These festivals are a high point of the social year for the villages and they are in fact the only times of feasting and dancing which are regularly scheduled on the calendar.

Anyone has an open invitation to attend a festival and people always come from many of the surrounding villages. Thus the frequency with which the people of one village attend the festival of other villages gives a rough measure of the amount of contact and interaction between the villages.

To determine the patterns of church festival attendance the following question was asked of the group of people assembled for an intelligibility test: "How often do people from your village attend the church festival at village X?" The responses to the question were not always reliable. In some cases the person answered that they went to all the festivals every time, but meant that they could go to any of them at any time if they wished. In some cases an individual would answer only for himself, instead of the village, telling how often he personally goes to the festivals. In the first case the answers were consistently too high; in the second they were consistently too low. In spite of attempts to rephrase the question, the proper kinds of response were not obtained in NEO, LWO, BAN, and NOP. Thus missing values (signified by periods) are reported for these four villages. The responses from Nanggu (NNG) look suspicious on first glance as they claimed that they attended all of the festivals at least some of the time. This claim is, however, consistent with their results on the intelligibility tests, their pattern of marriage ties, and their own opinions as to how well they understood the other dialects.

The results of the church festival question are set out in Table 2.8. The results are not strictly dialect to dialect contact; they are from central village of a dialect to central village. The list of villages on the left hand side of the table are the villages which were asked the question. The villages listed along the top are the villages where festivals are held. Thus, the first "2" in the second row of the table indicates that the people of Matu (MAT) attend the festival at Neo (NEO) every year.

Attraction and motivation are computed as they were for the opinions in the previous section. The attraction figures can be interpreted as the proportion of the island's population which attend the festivals at that location. It must be remembered, however, that attendance records for four of the dialects are not included in the sum of the attending population, but are still included in the total population figure. Thus the proportions are lower than they would be if comparable data from the four were added. The motivation figures can be interpreted as the proportion of the island's population which that group contacts in its festival attendance. These two sets of proportions should be qualified by stating that they do not apply to all individuals within the communities, but only to the delegations which represent them at festivals.

Table 2.8 Attendance at church festivals

"How often do people from your village attend the church festival at village X?"

2 = every year
1 = only some years
0 = never (or very seldom)

Village where festival is held:

Village asked:	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG	Mot
NEO	.	.	.	.	.	.	.	.	.	.	.	.	.	.
MAT	2	2	2	2	2	2	0	0	0	0	0	0	0	.55
BAN	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NEP	1	2	2	2	2	2	1	1	1	2	1	0	0	.70
LWO	.	.	.	.	.	.	.	.	.	.	.	.	.	.
VEN	2	2	2	2	2	2	2	2	1	1	1	0	0	.76
NEM	1	1	2	2	2	2	2	2	1	1	1	0	0	.72
BYO	1	1	1	1	2	2	2	2	2	2	1	0	0	.65
NOP	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NEA	0	0	2	2	1	1	1	2	2	2	2	1	1	.68
NOO	0	0	1	1	1	1	1	1	1	2	2	1	1	.48
MBI	0	0	0	0	0	0	1	1	0	1	2	2	2	.26
NNG	1	1	1	1	1	1	1	1	1	1	1	2	2	.52
Att	.29	.30	.55	.41	.53	.41	.36	.40	.35	.41	.34	.15	.13	

2.1.10 Contact through marriage ties

The present day network of marriage ties on Santa Cruz is set out in Table 2.9. At each of the thirteen intelligibility test points the people were asked how many people (either male or female) from their immediate dialect group were married to a person from each of the other dialects on the island. The answers to this question should produce reciprocal responses. That is, the people of MAT should answer the same number of marriages with BAN, as the people of BAN answer for marriages with MAT. Any discrepancies in the original data between the number of marriages as reported by different villages were rectified by assuming that the higher number was the correct number. This was done on the assumption that it was more likely that people would fail to think of a marriage with a particular dialect than that they would report one that was not really true.

The question, "How many people from your dialect are married to people from dialect X?" was scored as follows: 0 = none; 1 = one; 2 = some (two to four); 3 = many (five or more). When asking the question, the actual number of people was requested for the response. Sometimes, when many marriages were involved, the people were not able to think of every one and give an absolute number. This, combined with the fact of the discrepancies for which figures were adjusted and the different size of populations represented by the different dialects makes a scale of "none, one, some, many" preferable to the absolute numbers. The scale values which appear in Table 2.9 were assigned on the basis of the adjusted actual number of marriages reported.

Since the data in Table 2.9 are symmetric, measures of attraction and motivation cannot be computed. A better alternative to asking how many marriage ties link a pair of dialects, would have been to ask, "How many people from dialect X have married someone from here and are living here?" This would yield a nonsymmetric table of results. Since this question was not asked, the next best thing is to use the available data to predict what the results might be. To do this the following hypothesis is made: the number of couples residing in a particular village is proportional to the size of the village. Thus, if there are X number of marriage ties between two dialects with populations A and B, the number of those X couples living in dialect area A will be (X)(A/(A+B)), and the number of the couples living in dialect area B will be (X)(B/(A+B)).

In Table 2.10, the data in Table 2.9 are transformed as detailed above in order to reflect predicted patterns of marital residence. The dialects listed along the top are labeled place of residence, and those along the left hand side are labeled place of origin. The data are now nonsymmetric and measures of attraction and motivation can be computed. The row and column means are divided by three in order to compute a proportion from zero to one. The attraction figure can be loosely interpreted as the proportion of the island's population which has contact with that dialect because of marriage ties into that dialect. The motivation figure can be loosely interpreted as the proportion of the island's population with which the dialect has contact because of marriage ties outside that dialect.

2.1.11 Estimated intelligibility

In the field it was possible to test only 78 out of the possible 169 intelligibility relations among the 13 dialects. On the basis of the models developed in Section 6.3 to explain those 78 cases, the remaining untested relations can be estimated. Table 2.11 gives a complete matrix of estimated intelligibility. The estimates agree with the measurements in 95% of the cases. The four cases where

Table 2.9 Marriage ties

Number of marriage ties between dialects

0 = no marriages
1 = one marriage
2 = some marriages (two to four)
3 = many marriages (five or more)

	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
NEO	3	0	2	0	3	1	0	0	0	0	0	0	1
MAT	0	3	3	1	0	2	0	0	0	0	0	0	0
BAN	2	3	3	3	2	3	1	0	2	1	1	0	3
NEP	0	1	3	3	2	0	0	0	3	3	0	0	2
LWO	3	0	2	2	3	3	2	0	1	1	0	0	3
VEN	1	2	3	0	3	3	2	2	1	0	2	0	1
NEM	0	0	1	0	2	2	3	3	0	1	1	1	0
BYO	0	0	0	0	0	2	3	3	2	0	2	0	0
NOP	0	0	2	3	1	1	0	2	3	3	1	0	1
NEA	0	0	1	3	1	0	1	0	3	3	3	0	1
NOO	0	0	1	0	0	2	1	2	1	3	3	3	3
MBI	0	0	0	0	0	0	1	0	0	0	3	3	2
NNG	1	0	3	2	3	1	0	0	1	1	3	2	3

the estimate differs from the measurement are underlined.

Table 2.10 Predicted marital residence

Place of residence:

Place of Origin:	NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG	Mot
NEO	3.0	0.0	1.4	0.0	1.9	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.5	.19
MAT	0.0	3.0	2.4	0.7	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	.19
BAN	0.6	0.6	3.0	1.2	0.9	1.2	0.3	0.0	0.5	0.3	0.4	0.0	0.9	.22
NEP	0.0	0.3	1.8	3.0	1.1	0.0	0.0	0.0	0.9	1.2	0.0	0.0	0.8	.21
LWO	1.1	0.0	1.1	0.9	3.0	1.3	0.7	0.0	0.3	0.4	0.0	0.0	1.1	.22
VEN	0.4	0.6	1.8	0.0	1.7	3.0	0.8	0.8	0.3	0.0	1.0	0.0	0.4	.27
NEM	0.0	0.0	0.7	0.0	1.3	1.2	3.0	1.5	0.0	0.6	0.6	0.4	0.0	.21
BYO	0.0	0.0	0.0	0.0	0.0	1.2	1.5	3.0	0.9	0.0	1.2	0.0	0.0	.12
NOP	0.0	0.0	1.5	2.1	0.7	0.7	0.0	1.1	3.0	1.8	0.7	0.0	0.6	.31
NEA	0.0	0.0	0.7	1.8	0.6	0.0	0.4	0.0	1.2	0.0	1.7	0.0	0.5	.22
NOO	0.0	0.0	0.6	0.0	0.0	1.0	0.4	0.8	0.3	1.3	3.0	1.0	1.2	.18
MBI	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	2.0	3.0	1.2	.10
NNG	0.5	0.0	2.1	1.2	1.9	0.6	0.0	0.0	0.4	0.5	1.7	0.8	3.0	.36
Att	.10	.06	.40	.23	.31	.27	.13	.10	.15	.18	.22	.06	.23	

The method used to estimate intelligibility was a two-out-of-three method for combining the three best predicting models. In most cases, the three models agree. In the cases where they do not, the level predicted by two of the models is taken as the estimated intelligibility.

The first model is based on composite relative distance alone (Section 6.4) where composite relative distance equals six-tenths times relative geographic distance (Table 2.2) plus four-tenths times relative lexical distance (Table 2.5). The step function for predicting intelligibility is (see final scattergram in Appendix 2.2),

Int = 3, if composite distance $\leq$ 134%;
= 2, if 134% $<$ composite distance $\leq$ 185%;
= 1, if 185% $<$ composite distance.

This model is 90% accurate.

The second model is a complex model (Section 6.5) with predicted contact. The contact factor is predicted by the overall motivation of the listener's dialect as indicated by opinions about intelligibility (Table 2.7) divided by the relative geographic distance from the listener's dialect to the speaker's (Table 2.2). The scaling factors for these two variables are described in Appendix 2.3. After the two variables are scaled, they are multiplied to compute the C factor which plugs into the formula for familiarity,

$$F = L + C(100-L)$$

The step function which predicts intelligibility is then,

Int = 3, if 89% $<$ Familiarity $\leq$ 100%;
= 2, if 82% $\leq$ Familiarity $\leq$ 89%;
= 1, if Familiarity $<$ 82%.

This model is 92 % accurate.

The third model is also a complex model with predicted contact. Lexical distance from the center is used to estimate attraction and motivation. Contact is predicted by the attraction of the speaker (inverse of distance from center) times the motivation of the hearer (distance from center) divided by relative geographic distance. The scaling factors for these three variables are described in Appendix 2.3. Contact is plugged into the familiarity formula as above. The step function which predicts intelligibility is,

Int = 3, if 75% $<$ Familiarity $\leq$ 100%;
= 2, if 64% $<$ Familiarity $\leq$ 75%;
= 1, if Familiarity $\leq$ 64%.

This model is 90% accurate.

Table 2.11 Estimated intelligibility

3 = Full intelligibility
2 = Partial intelligibility
1 = Sporadic recognition

Dialect of speaker:

		NEO	MAT	BAN	NEP	LWO	VEN	NEM	BYO	NOP	NEA	NOO	MBI	NNG
Dialect of hearers:	NEO	3	3	3	3	3	3	3	3	2	2	2	1	1
	MAT	3	3	3	3	3	3	3	2	2	2	1	1	1
	BAN	3	3	3	3	3	3	3	2	2	2	2	1	1
	NEP	3	3	3	3	3	3	3	3	3	3	2	1	1
	LWO	3	3	3	3	3	3	3	3	2	2	2	1	1
	VEN	3	3	3	3	3	3	3	3	2	2	2	1	1
	NEM	3	3	3	3	3	3	3	3	3	3	2	1	1
	BYO	3	3	3	3	3	3	3	3	3	3	3	1	1
	NOP	2	2	3	3	3	3	3	3	3	3	3	2	1
	NEA	2	2	3	3	3	3	3	3	3	3	3	3	1
	NOO	2	2	3	3	3	3	3	3	3	3	3	3	3
	MBI	2	2	3	3	3	3	3	3	3	3	3	3	3
	NNG	3	3	3	3	3	3	3	3	3	3	3	3	3

Int = 3, if composite distance ≤ 134%;

= 2, if 134% < composite distance ≤ 185%;

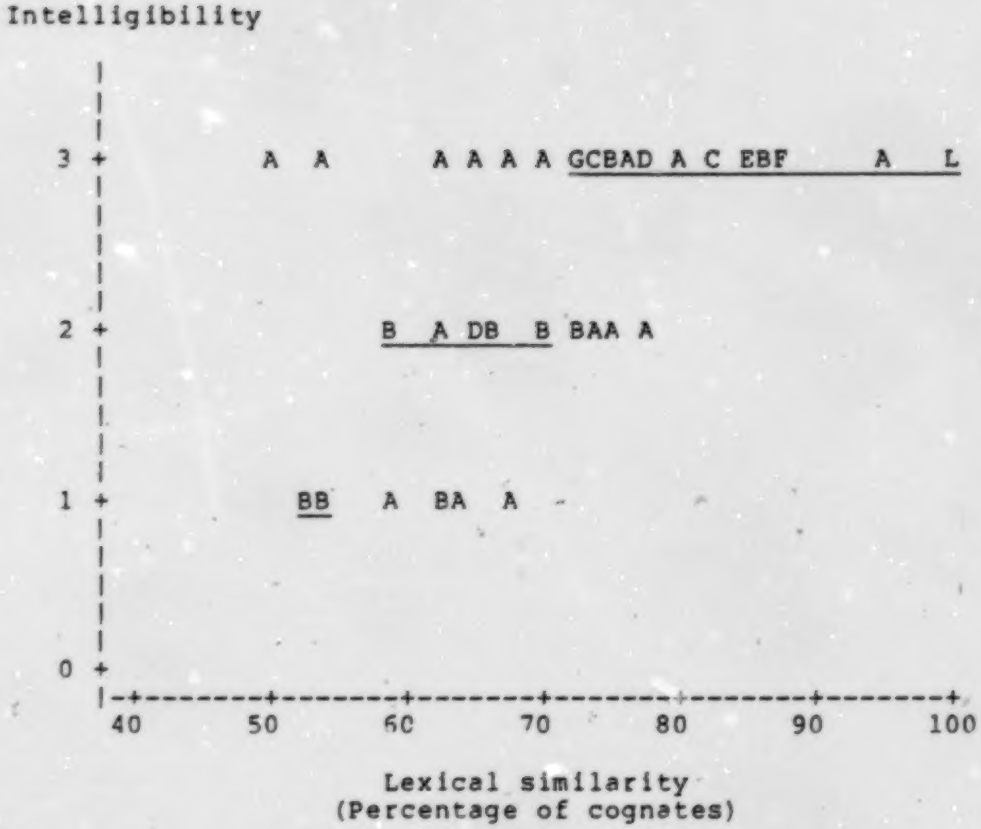
= 1, if 185% < composite distance.

This model is 90% accurate.

2.3 Scattergrams and step functions for single variable models

In the scattergrams, intelligibility is plotted on the vertical axis and the predicting variable is plotted on the horizontal axis. The plotted values are the letters of the alphabet. A indicates that one observation is plotted at that point, B indicates that two are, and so on. The steps of the step functions are indicated by underscores. Below each scattergram three values are given: the sum of the deviations of predicted values of intelligibility from the measured values, the ratio of prediction accuracy, and the percentage of prediction accuracy.

Lexical Similarity



Sum of Deviations = 18

Ratio of Accuracy = 60/78

Percentage of Acc = 77%

Absolute Geographic Distance

```
Intelligibility

  |
  |
  |
3 + DFD  DABDDAABCC ADA B A  A       A A A  A
  |
  |
  |
  |
  |
2 +           A AC ABC A B        A        A
  |
  |
  |
  |
  |
1 +                       A   A    AAA  AA  A A
  |
  |
  |
  |
  |
0 +
  |-+---------+---------+---------+---------+---------+
    0        200       400       600       800       1000
```

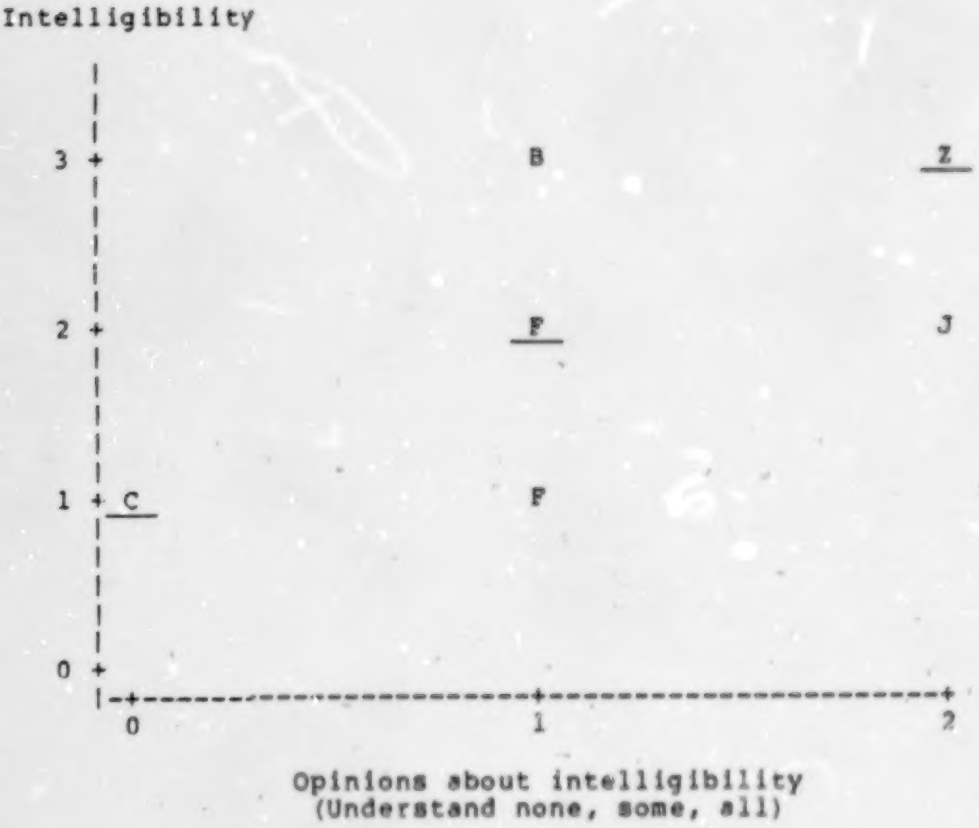
Geographic distance
(Minutes' traveling time)

Sum of Deviations = 26

Ratio of Accuracy = 52/78

Percentage of Acc = 67%

Opinions about Intelligibility

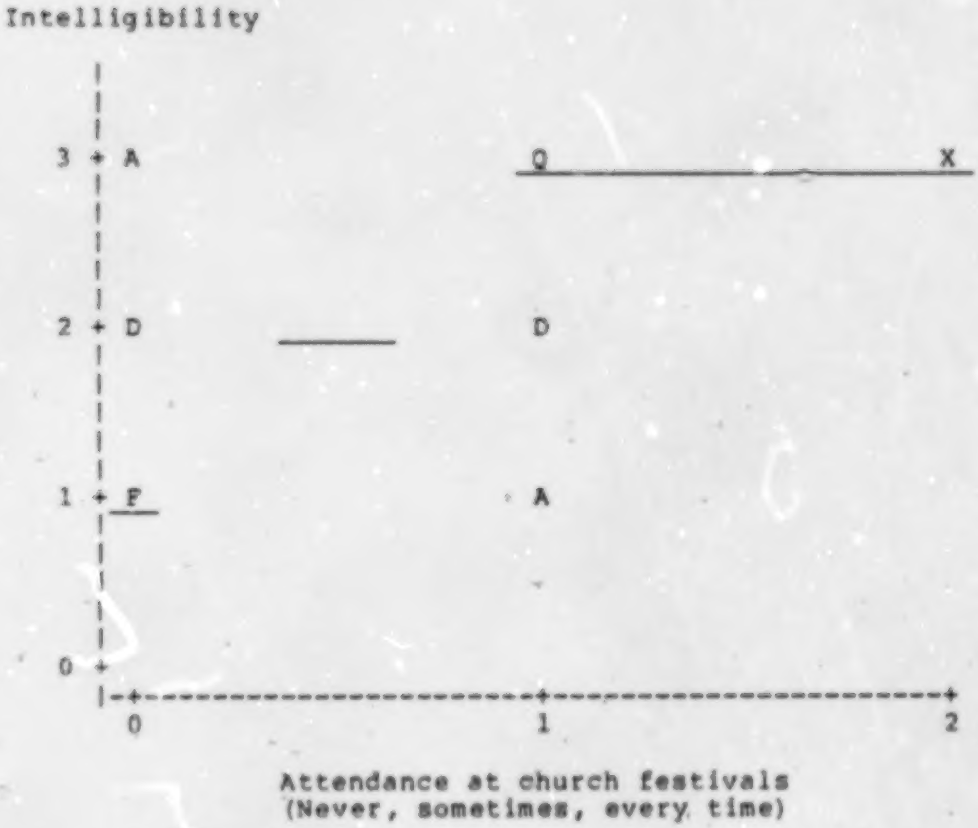


Sum of Deviations = 18

Ratio of Accuracy = 60/78

Percentage of Acc = 77%

Attendance at Church Festivals

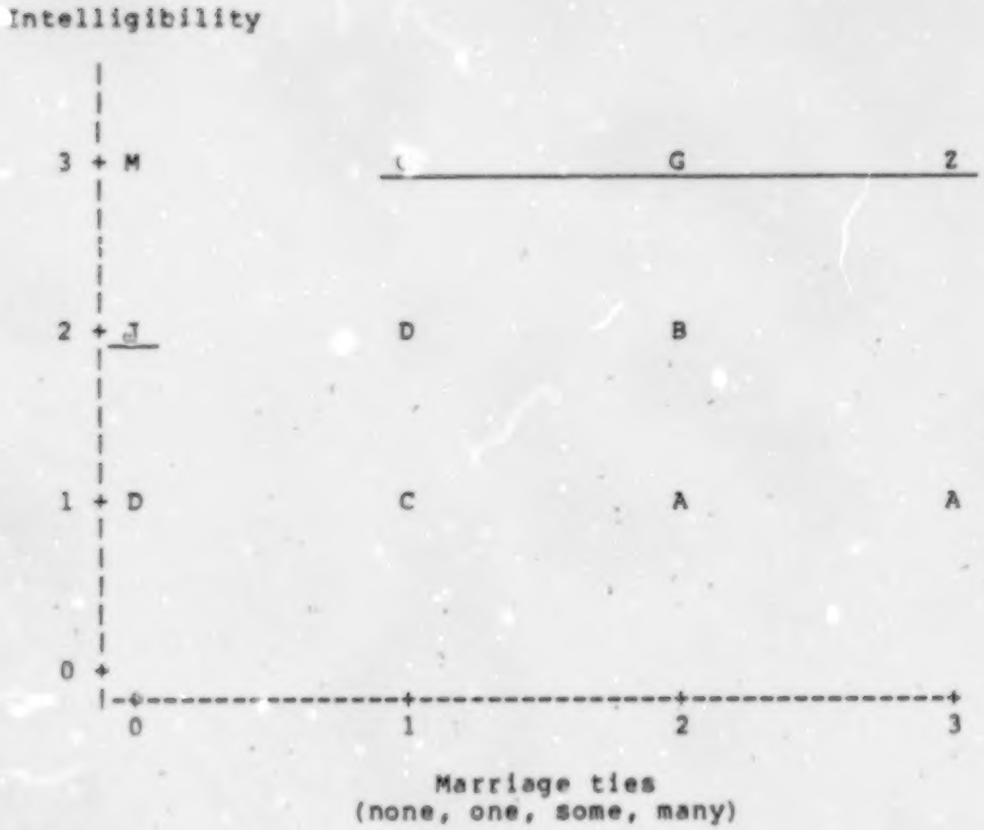


Sum of Deviations = 12

Ratio of Accuracy = 45/57

Percentage of Acc = 80%

Marriage Ties

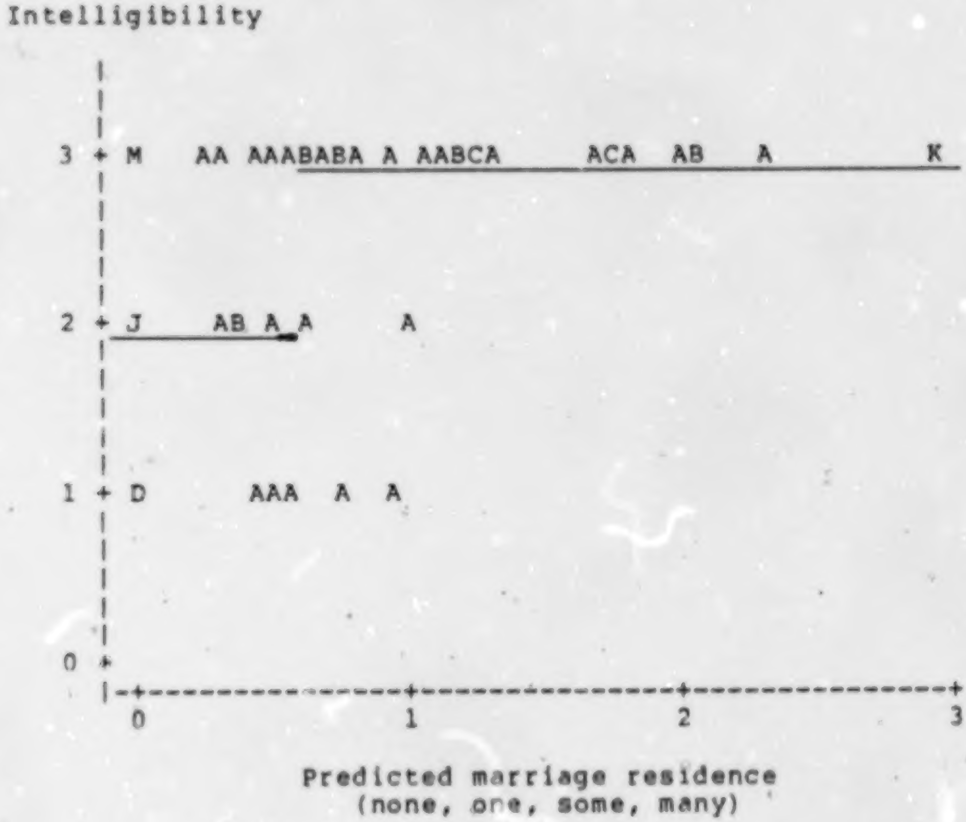


Sum of Deviations = 33

Ratio of Accuracy = 45/78

Percentage of Acc = 58%

Predicted Marriage Residence

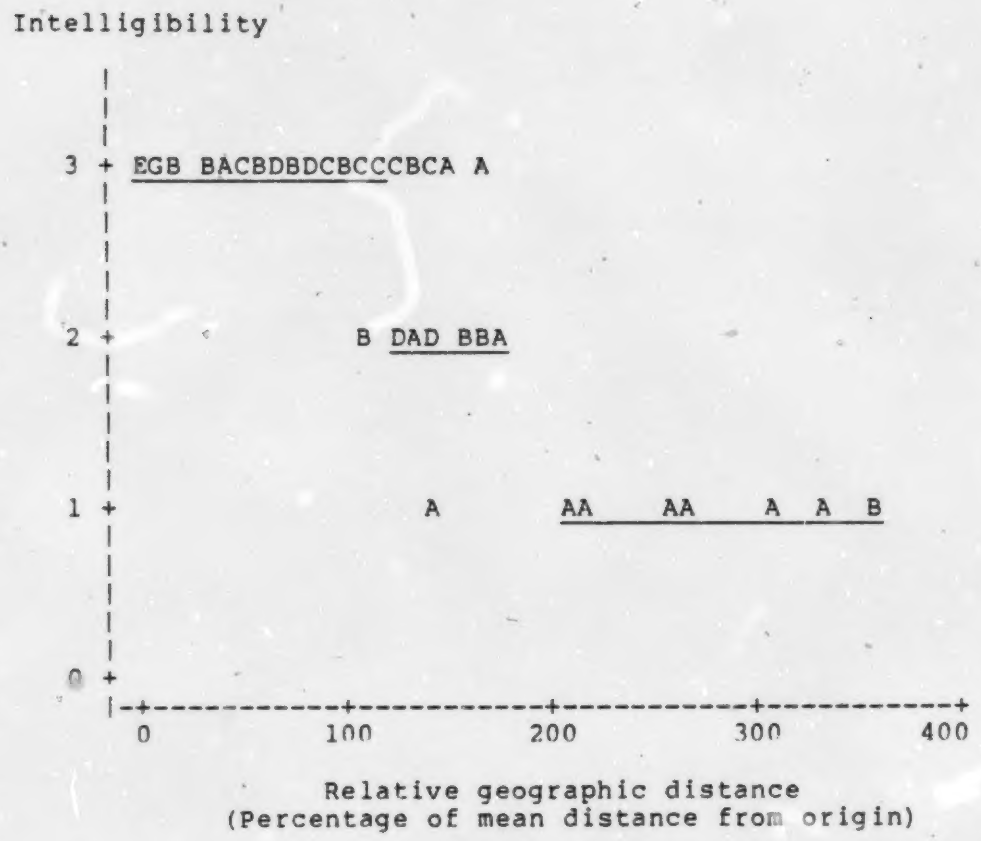


Sum of Deviations = 31

Ratio of Accuracy = 47/78

Percentage of Acc = 60%

Relative Geographic Distance



Sum of Deviations = 13

Ratio of Accuracy = 65/78

Percentage of Acc = 83%

Relative Lexical Distance

```
Intelligibility

     |
     |
     |
   3 + L   A   BBAADDDCD ACADCA  B
     |
     |
     |
     |
     |
   2 +                     A B DB ABA   AAA
     |
     |
     |
     |
     |
   1 +                       A  AA      AA    A  AAA
     |
     |
     |
     |
     |
   0 +
     |-+-----------+-----------+-----------+-----------+
       0          100         200         300         400

                   Relative lexical distance
            (Percentage of mean distance from origin)
```

Sum of Deviations = 11

Ratio of Accuracy = 67/78

Percentage of Acc = 86%

Composite Relative Distance

```
Intelligibility

      |
      |
      |
   3 + GD    BEBABBACBCACDCAC BA  A
      |
      |
      |
      |
      |
   2 +                   A  BBE AAAAB
      |
      |
      |
      |
      |
   1 +                            A  B    A        AA    A   B
      |
      |
      |
      |
      |
   0 +
      |-+--------------+--------------+--------------+-----
        0             100            200            300

              Composite relative distance
         (Percentage of mean distance from origin)
```

Sum of Deviations = 8

Ratio of Accuracy = 70/78

Percentage of Acc = 90%

2.3 Scaling of variables for inclusion in contact factor

The contact factor in the predicting equation must take on the range of zero to one to prevent predicting more than 100% intelligibility. The variables are scaled by the following formula:

scaled value = (raw value - min) / (max - min)

Min is the value for that variable which should scale to zero, max is the value which should scale to one. Scaled values less than zero are set to zero, and those greater than one are set to one. Note that when the min value is zero, the formula reduces to a simple division: raw value/max. In the case of the measured contact models, the raw values are divided by the max values listed below and plugged straight into the prediction formula. In the case of the predicted contact models, as many as three variables are involved: attraction, motivation, and distance. In the case of opinions, festivals, and marriages, the raw values for attraction and motivation are take from the outer row and column of the data tables in Appendix 2.1. In the case of the other four factors, the raw values are population, density, or distance from the center (Mbanua). In all cases, the attraction measure is associated with the speaker and the motivation measure is associated with the hearer. The third variable involved in predictions is distance. The distance measures are scaled so as to invert them, that is, far distance yields a low value and close distance yields a high value. In this way the three scaled variables can be multiplied to compute a contact factor in the scale of zero to one. The scaling of population and density was handled a little differently. The equations used are reported in the following table. Also, for these two variables, the product of attraction and motivation was further scaled as indicated in the table. With all of this information it should be possible to replicate the results I report in the remaining sections of this appendix.

Measured contact	min	max
opinions	0	2
festivals	0	2
marriage ties	0	3
marriage residence	0	3

Predicted contact	min	max
absolute geographic distance	830	105
relative geographic distance	361	40
absolute lexical distance	50	5
relative lexical distance	290	30
opinions, attraction	0	1
opinions, motivation	0	1
festivals, attraction	0	.55
festivals, motivation	0	.76
marriages, attraction	0	.39
marriages, motivation	0	.35
geographic distance from center, attraction	650	60
geographic distance from center, motivation	60	650
lexical distance from center, attraction	46	5
lexical distance from center, motivation	5	46
population, attraction	raw/450	
population, motivation	120/raw	
population, attraction x motivation	(attr)(mot)/.67	
density, attraction	raw/520	
density, motivation	129/raw	
density, attraction x motivation	(attr)(mot)/.67	

2.5 Scattergrams for complex models with measured contact

Opinions as measure of contact

```
Intelligibility

  |
  |
  |
3 +                                     A      A      Z
  |
  |
  |
  |
  |
2 +                               A  A  AAAA          J
  |
  |
  |
  |
  |
1 +    AA        A             B  A B A
  |
  |
  |
  |
  |
0 +
  |-+---------+---------+---------+---------+---------+
   50        60        70        80        90        100

                       Familiarity
```

Sum of Deviations = 13

Ratio of Accuracy = 65/78

Percentage of Acc = 83%

Festival attendance as measure of contact

Intelligibility

```
  |
  |
  |
3 +                          A A    AA BCCBA  BA      X
  |                                     _______________
  |
  |
  |
  |
2 +          A     AA        A      AA B
  |                ______________________
  |
  |
  |
  |
1 +    BA    A   B                     A
  |    ___________
  |
  |
  |
  |
0 +
  |-+---------+---------+---------+---------+---------+
   50        60        70        80        90        100
```

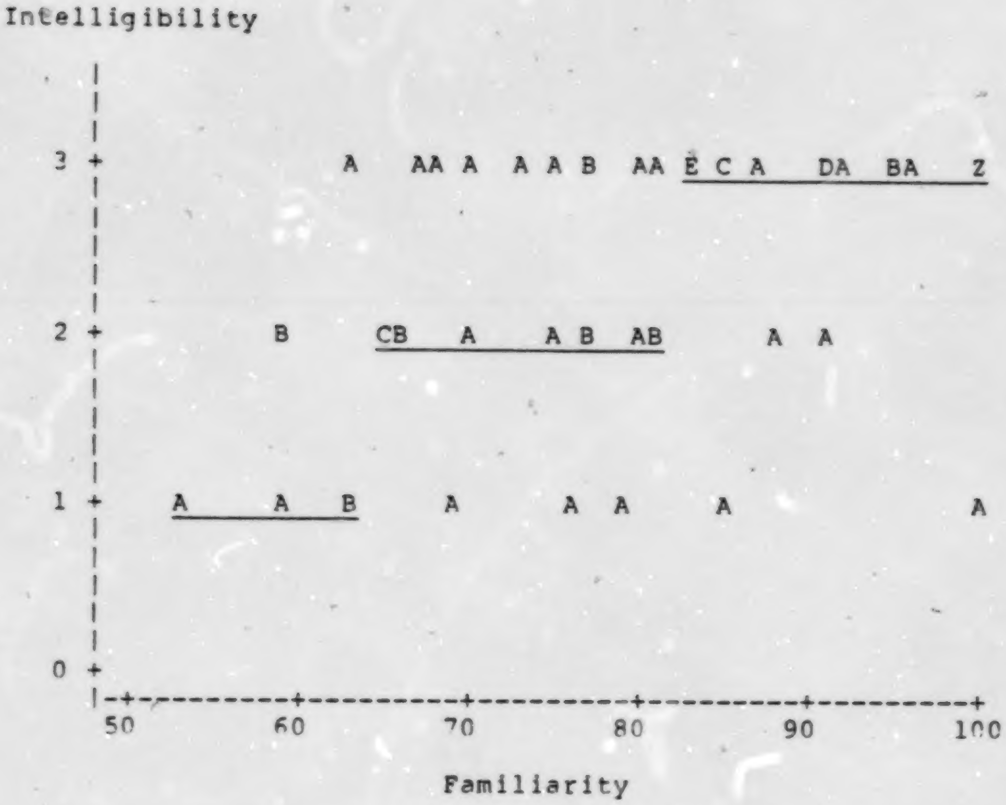
Familiarity

Sum of Deviations = 8

Ratio of Accuracy = 49/57

Percentage of Acc = 85%

Marriage ties as measure of contact



Sum of Deviations = 20

Ratio of Accuracy = 58/78

Percentage of Acc = 74%

Marriage residence as measure of contact

```
Intelligibility
  |
  |
  |
3 +         A    A    A A AA BAA  C AEBDAACAA CC B   AL
  |
  |
  |
  |
  |
2 +          B     CB  AA    DAAA
  |
  |
  |
  |
  |
1 +    A     B   B  A A  A A
  |
  |
  |
  |
  |
0 +
  |-+---------+---------+---------+---------+---------+
   50        60        70        80        90       100

                    Familiarity
```

Sum of Deviations = 18

Ratio of Accuracy = 60/78

Percentage of Acc = 77%

2.5 Results for complex models with predicted contact

Following are seven tables, one for each of the variables used to predict contact. For each variable, eighteen different combinations of values in the numerator and denominator of the contact formula were used. The three rows represent three different numerators: attraction alone, motivation alone, and attraction times motivation. The six columns represent six different denominators: no distance (a constant value of one), absolute geographic distance, relative geographic distance, absolute lexical distance, relative lexical distance, and composite relative distance. At the intersection of each row and column two values are given. The first is the sum of the deviations of predicted from measured values of intelligibility; the second, in parentheses, is the percentage of prediction accuracy. The total number of predictions on which the percentages are based is given in the heading for each table.

Opinions about intelligibility
(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	13 (83%)	20 (74%)	12 (85%)	18 (77%)	9 (88%)	11 (86%)
Motivation	18 (77%)	16 (79%)	6 (92%)	14 (82%)	8 (90%)	7 (91%)
Attr & Mot	10 (87%)	16 (79%)	8 (90%)	14 (82%)	6 (92%)	6 (92%)

Church festival attendance
(57 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	8 (86%)	12 (79%)	8 (86%)	11 (81%)	7 (88%)	9 (84%)
Motivation	20 (65%)	17 (70%)	12 (79%)	15 (74%)	12 (79%)	12 (79%)
Attr & Mot	13 (77%)	12 (79%)	10 (82%)	10 (82%)	8 (86%)	8 (86%)

Marriage residence
(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	23 (71%)	19 (76%)	16 (79%)	17 (78%)	16 (79%)	16 (79%)
Motivation	25 (68%)	21 (73%)	15 (81%)	18 (77%)	11 (86%)	14 (82%)
Attr & Mot	22 (72%)	19 (76%)	16 (79%)	17 (78%)	16 (79%)	16 (79%)

Population

(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	23 (71%)	17 (78%)	16 (79%)	18 (77%)	15 (81%)	16 (79%)
Motivation	22 (72%)	16 (79%)	13 (83%)	16 (79%)	14 (82%)	14 (82%)
Attr & Mot	19 (76%)	16 (79%)	15 (81%)	18 (77%)	16 (79%)	15 (81%)

Density of population

(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	21 (73%)	18 (77%)	16 (79%)	16 (79%)	14 (82%)	15 (81%)
Motivation	22 (72%)	15 (81%)	15 (81%)	15 (81%)	14 (82%)	15 (81%)
Attr & Mot	18 (77%)	15 (81%)	13 (83%)	17 (78%)	12 (85%)	14 (82%)

Geographic distance from center

(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	16 (79%)	19 (76%)	13 (83%)	17 (78%)	10 (87%)	12 (85%)
Motivation	14 (82%)	13 (83%)	10 (87%)	14 (82%)	11 (86%)	10 (87%)
Attr & Mot	11 (86%)	15 (81%)	11 (86%)	15 (81%)	10 (87%)	11 (86%)

Lexical distance from center

(78 total predictions)

	None	Geographic		Lexical		Composite
		Absolute	Relative	Absolute	Relative	Relative
Attraction	14 (82%)	18 (77%)	12 (85%)	15 (81%)	11 (86%)	12 (85%)
Motivation	15 (81%)	13 (83%)	10 (87%)	13 (83%)	8 (90%)	10 (87%)
Attr & Mot	11 (86%)	14 (82%)	8 (90%)	13 (83%)	9 (88%)	8 (90%)

BIBLIOGRAPHY

Ascher, Marcia and Robert Ascher. 1963. Chronological ordering by computer. American Anthropologist 65:1045-52.

Bender, Marvin L. 1971. The languages of Ethiopia: a new lexicostatistic classification and some problems of diffusion. Anthropological Linguistics 13:165-288.

----- and Robert L. Cooper. 1971. Mutual intelligibility within Sidamo. Lingua 27:32-52.

Berlin, Brent and Paul Kay. 1969. Basic color terms: their universality and evolution. Berkeley: University of California Press.

Biggs, Bruce. 1957. Testing intelligibility among Yuman languages. IJAL 23:57-62.

Blalock, Hobert M., Jr. 1972. Social statistics, second edition. New York: McGraw-Hill.

Bloomfield, Leonard. 1933. Language. New York: Henry Holt and company.

Bradley, C. Henry. 1968. A method for determining dialectal boundaries and relationships. America Indigena 28:751-760.

Callister, Sandra. 1977. Sociolinguistic approaches to dialect surveying in Papua New Guinea. In Loving and Simons 1977, pp. 187-216.

Capell, Arthur. 1962. The techniques of structural statistics. Oceania 33:1-11.

Carrothers, Gerald A. P. 1956. An historical review of the gravity and potential concepts of human interaction. Journal of the American Institute of Planners 22:94-102.

Casad, Eugene H. 1974. Dialect intelligibility testing. Summer Institute of Linguistics Publications in Linguistics and Related Fields, Number 38. Norman, Oklahoma: Summer Institute of Linguistics of the University of Oklahoma.

Collier, Ken J. 1977. Predicting intelligibility: a suggested technique. In Loving and Simons 1977, pp. 253-262.

Combs, Martin. 1977. Cultural considerations in language change and communication. In Loving and Simod Simons 1977, pp. 217-232.

Darlington, Richard. 1975. Radicals and squares and other statistical procedures for the behaviorial sciences. Ithaca, NY: Logan Hill Press.

de Saussure, Ferdinand. 1959. Course in general linguistics. Translated by Wade Baskin. New York: Philosophical Library.

Downie, N. M. and R. W. Heath. 1974. Basic statistical methods, fourth edition. New York: Harper and Row.

Elbert, Samuel H. 1953. Internal relationships of Polynesian languages and dialects. Southwestern Journal of Anthropology 9:147-173.

Ellegård, Alvar. 1959. Statistical measurement of linguistic relationship. Language 35:131-156.

Fodor, István. 1965. The rate of linguistic change: limits of the application of mathematical methods in linguistics. Janua Linguarum, series minor, number 43. The Hague: Mouton.

Glasgow, David and Richard Loving. 1964. Languages of the Maprik sub-district. Port Moresby, Papua New Guinea: Department of Information and Extension Services.

Gleason, H. A. 1959. Counting and calculating for historical reconstruction. Anthropological Linguistics 1.2:22-32.

Greenberg, Joseph H. 1960. A quantitative approach to the morphological typology of language. IJAL 26:178-194.

Grimes, Joseph E. 1964. Measures of linguistic divergence. In Proceedings of the ninth international congress of linguists, edited by H. G. Lunt. The Hague: Mouton, pp. 44-50.

-----. 1974. Dialects as optimal communication networks. Language 50:260-269.

----- and Frederick Agard. 1959. Linguistic divergence in Romance. Language 35:598-604.

Gudschinsky, Sarah C. 1955. Lexicostatistic skewing from dialect borrowing. IJAL 21:138-149.

-----. 1956. The ABC's of lexicostatistics (glottochronology). Word 12:175-210.

Halliday, M. A. K. 1961. Categories of the theory of grammar. Word 17:241-292.

Harbeck, Warren A. 1969. A study in mutual intelligibility and linguistic separation among five Siouan languages. ms.

Hickerson, H., Glen D. Turner, and Nancy P. Hickerson. 1952. Testing procedures for estimating transfer of information among Iroquois dialects and languages. IJAL 18:1-8.

Hockett, Charles F. 1958. A course in modern linguistics. New York: Macmillan.

Hymes, Dell H. 1960. Lexicostatistics so far. Current Anthropology 1:3-44.

Ivič, Pavle. 1974. Dialects. Encyclopaedia Britannica, Macropaedia, 5:696-701.

Kroeber, A. L. 1960. On typological indices I: ranking of langauges. IJAL 26:171-177.

----- and C. D. Chrétien. 1937. Quantitative classification of Indo-European languages. Language 13:83-103.

----- and -----. 1939. The statistical technique and Hittite. Language 15:69-71.

Ladefoged, Peter. 1968. The measurement of cross-language communication. Uganda Language Survey Working Paper. Educational Resources Information Center Document ED 024 920 (AL 001 284). Washington: U. S. Office of Health, Education, and Welfare.

-----. 1970. The measurement of phonetic similarity. Statistical Methods in Linguistics 6:23-32.

-----, Ruth Glick, and Clive Criper. 1972. Language in Uganda. London: Oxford University Press.

Lees, Robert B. 1953. The basis of glottochronology. Language 29:113-127.

Lounsbury, Floyd G. 1961. Iroquois-Cherokee linguistic relations. Smithsonian Institution Bureau of American Ethnology Bulletin 180:11-17.

Loving, Richard and Gary F. Simons, eds. 1977. Language variation and survey techniques. Workpapers in Papua New Guinea languages, volume 21. Ukarumpa, E. H. P., Papua New Guinea: Summer Institute of Linguistics.

Malinowski, Bronislaw. 1922. Argonauts of the western Pacific. London: Routledge and Kegan Paul.

McElhanon, K. A. 1970. Lexicostatistics and the classification of Huon Peninsula languages. Oceania 40:214-31.

-----. 1971. Classifying New Guinea languages. Anthropos 66:120-144.

McKaughan, Howard. 1964. A study of divergence in four New Guinea languages. American Anthropologist 66.4.2:98-120.

Miller, Delbert C. 1977. Handbook of research design and social measurement, third edition. New York: David McKay Co.

Moore, Bruce R. 1961. A statistical morpho-syntactic typology study of Colorado (Chibcha). IJAL 27:298-307.

Nadel, S. F. 1957. The theory of social structure. Glencoe, Illinois: The Free Press.

Olmsted, David L. 1954a. Achumawi-Atsugewi non-reciprocal intelligibility. IJAL 20:181-184.

-----. 1954b. Towards a cultural theory of lexical innovation: a research design. In Report of the fifth annual round table meeting on linguistics and language teaching, ed. by Hugo Mueller. Georgetown University Monograph Series on

Languages and Linguistics 7:105-117.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum. 1957. The measurement of meaning. Urbana: University of Illinois Press.

Pierce, Joe E. 1952. Dialect distance testing in Algonquian. IJAL 18:203-210.

Pike, Kenneth L. 1967. Language in relation to a unified theory of the structure of human behavior. Second, revised edition. The Hague: Mouton.

----- and Evelyn G. Pike. 1977. Grammatical analysis. Summer Institute of Linguistics Publications in Linguistics, Number 53. Dallas, Texas: Summer Institute of Linguistics and the University of Texas at Arlington.

Rae, John A. 1958. Concerning the validity of lexicostatistics. IJAL 27:145-150.

Sanders, Arden. 1977a. Guidelines for conducting a lexicostatistic survey in Papua New Guinea. In Loving and Simons 1977, pp. 21-41.

-----. 1977b. Some synchronic analysis procedures for language survey data. In Loving and Simons 1977, pp. 295-316.

Sanders, Joy. 1977. On defining the center of a linguistic group. In Loving and Simons 1977, pp. 263-294.

Sankoff, Gillian. 1968. Social aspects of multilingualism in New Guinea. McGill University Ph.D. dissertation.

-----. 1969. Mutual intelligibility, bilingualism and linguistic boundaries. In International Days of Sociolinguistics. Rome: Instituto Luigi Sturzo. pp. 839-848.

Simons, Gary F. 1977a. A dialect survey of Santa Cruz Island. Working papers for the language variation and limits to communication project, number 3. Ithaca, NY: Cornell University.

-----. 1977b. A package of computer programs for the analysis of language survey word lists. Working papers for the language variation and limits to communication project, number 7. Ithaca, NY: Cornell University.

-----. 1977c. Phonostatistic methods. In Loving and Simons 1977, pp. 155-186.

-----. 1977d. The role of purpose and perspective in planning a language survey. In Loving and Simons 1977, pp. 9-20.

Simons, Linda L. 1977. Guidelines for intelligibility testing in Papua New Guinea. In Loving and Simons 1977, pp. 233-251.

Snider, James G. and Charles E. Osgood. 1969. Semantic differential technique: a sourcebook. Chicago: Aldine.

St. Clair, Robert. 1974a, Communication across linguistically related systems. Linguistics 126:83-98.

-----. 1974b. Communication across Polynesian languages. Kivung 7:133-150.

Stevens, S. S. 1946. On the theory of scales of measurement. Science 103:677-680.

Stewart, J. Q. 1941. An inverse distance variation for certain social influences. Science 93:89-90.

-----. 1947a. Empirical mathematical rules concerning the distribution and equilibrium of population. Geographical review 37:461-85.

-----. 1947b. Suggested principles of "social physics". Science 106:179-80.

-----. 1952. A basis for social physics. Impact of science on society 3:110-33.

Stolzfus, Ronald D. 1974. Toward defining centers for indigenous literature programs: a problem of language communication. Cornell University M.A. thesis.

Swadesh, Morris. 1950. Salish internal relationships. IJAL 16:157-167.

-----. 1952. Lexicostatistic dating of prehistoric ethnic contacts. Proceedings of the American philosophical society 96:452-63.

-----. 1955. Towards greater accuracy in lexicostatistic dating. IJAL 21:121-132.

Voegelin, C. F. 1961. Typology of density ranges II: contrastive and non-contrastive syntax. IJAL 27:287-297.

----- and Zellig S. Harris. 1951. Methods for determining intelligibility among dialects of natural languages. Proceedings of the American Philosophical

Society 95:322-329.

-----, A. K. Ramanujan, and F. M. Voegelin. 1960. Typology of density ranges I: introduction. IJAL 26:198-205.

Ward, Jack H. 1962. Mutual intelligibility between certain Polynesian speech communities. University of Hawaii M.A. thesis.

Whiteman, Darrell and Gary Simons. 1977. The languages of Santa Isabel: a sociolinguistic survey. ms.

Winter, Werner. 1957. Yuman languages I: first impressions. IJAL 23:18-23.

Wolff, Hans. 1959. Intelligibility and inter-ethnic attitudes. Anthropological Linguistics 1.3:34-41.

Wurm, S. A. and D. C. Laycock. 1961. The question of language and dialect in New Guinea. Oceania 32:128-143.

Yamagiwa, Joseph K. 1967. On dialect intelligibility in Japan. Anthropological Linguistics 9(1):1-17.